Université Paris Cité

École doctorale 622 - Sciences du langage
Approches Linguistiques Théoriques, Appliquées et Expérimentales : langues et cultures connectées (URP 3967)

Articulatory and sociocultural constraints in the singing voice

The case of intervocalic /t/ in 1980s British popular music

Par Coline CAILLOL

Thèse de doctorat de Linguistique Anglaise

Dirigée par Emmanuel FERRAGNE

Présentée et soutenue publiquement le 14/11/2025

Devant un jury composé de :

Claire PILLOT-LOISEAU, PU	Université Sorbonne Nouvelle	Rapportrice
Jane STUART-SMITH, Professor	University of Glasgow	Rapportrice
Joan BEAL, Professor emeritus	University of Sheffield	Examinatrice
Sylvain NAVARRO, MCF	Université Paris Cité	Examinateur
Emmanuel FERRAGNE, PU	Université Paris Cité	Directeur de thèse

Abstract

Titre: Contraintes phonétiques articulatoires et socio-culturelles dans la voix chantée : le cas du /t/ intervocalique dans la musique populaire britannique des années 1980.

Mots clefs: sociophonétique ; phonétique articulatoire ; phonétique experimentale; voix chantée ; popular music studies ; electropalatographie (EPG) ; IRM ; prononciation du /t/

Résumé: À la croisée de la sociophonétique, des études sur les musiques populaires et de la phonétique expérimentale, cette thèse vise à étudier un phénomène désormais bien identifié dans le domaine de la recherche sur la voix chantée : la tendance des artistes britanniques de musique populaire à adopter une prononciation influencée par un accent américain lorsqu'ils chantent, mais pas lorsqu'ils parlent. Le premier volet, sociophonétique, de cette thèse propose une analyse quantitative des productions musicales britanniques (notamment dans le heavy metal traditionnel et la pop) des années 1980, une décennie marquante pour la création et la commercialisation de la musique populaire. L'étendue de cette modification de l'accent sera étudiée, avec une attention particulière à l'articulation du /t/ intervocalique. En tant que variable sociolinguistique, la réalisation de ce trait phonétique sous forme de battue (plus proche à l'oreille d'un [d] que d'un [t] occlusif prototypique) constitue à la fois un indice d'identité américaine, ainsi qu'un marqueur de conformité à des normes de prononciation supra-locales standardisées dans la musique populaire et connues sous le nom de *Pop Song English*. Dans cette première partie, des facteurs d'ordre socioculturel–dont la domination américaine de l'industrie musicale, l'importance des influences musicales et les expressions d'identités conflictuelles qui en résultent–sont examinés. Une analyse acoustique d'autres variables phonétiques (les voyelles de TRAP et LOT), ainsi qu'une étude sur l'évolution de la perception du /t/ auprès de locuteurs britanniques viennent compléter l'ensemble et l'inscrire dans une approche globale. Le deuxième volet, en phonétique articulatoire, explore une potentielle cause complémentaire menant à la réalisation du /t/ intervocalique en battue chez des artistes britanniques : les contraintes phonétiques qui distinguent la parole de la voix chantée. Une approche expérimentale novatrice dans l'étude de la musique pop-

ulaire utilisant plusieurs méthodes d'imagerie dont le lip tracking, l'électropalatographie (EPG) et l'imagerie par résonance magnétique (IRM) est proposée avec la participation de cinq chanteurs britanniques professionnels. Ces techniques permettent d'étudier successivement l'ouverture labiale, le contact de la langue avec le palais, et la configuration générale du conduit vocal, afin de tester l'influence de facteurs tels que la hauteur de la voix et la durée des phonèmes sur l'articulation dans le chant. L'hypothèse sous-jacente est que plus les conditions sont extrêmes (e.g. recours au registre aigu caractéristique du heavy metal), moins elles favorisent une occlusion nette de la langue contre la zone alvéolaire, et plus la réalisation du /t/ tend vers une prononciation en battue. Ainsi, la modification linguistique observée entre la parole et le chant découle peut-être également d'une stratégie de simplification articulatoire lors d'une activité complexe à bien des égards. Les résultats montrent que l'adoption par les chanteurs britanniques des années 1980 du /t/ battu dans le chant s'explique à la fois par une orientation socio-indexicale vers le modèle américain et les normes de prononciation de la pop, ainsi que par des ajustements articulatoires propres au chant. Ils révèlent également que, dans ce cadre, la variation individuelle, aussi subtile soit-elle sur le plan articulatoire, demeure marquée, qu'elle s'exprime à travers des attitudes personnelles ou par des stratégies physiologiques spécifiques.

Title: Articulatory and sociocultural constraints in the singing voice: the case of intervocalic /t/ in 1980s British popular music.

Keywords: sociophonetics; articulatory phonetics; experimental phonetics; singing voice, popular music studies; electropalatography (EPG); MRI; /t/ flapping

Abstract: At the intersection of sociophonetics, popular music studies, and experimental phonetics, this thesis aims to examine a now well documented phenomenon in research on the singing voice: the tendency for British popular music singers to adopt a pronunciation influenced by an American accent when singing, which does not occur when speaking. The first component of this work, sociophonetic in nature, presents a quantitative analysis of music productions (specifically from traditional British heavy metal and pop) of the 1980s, a pivotal decade for the creation and commercialization of popular music. The extent of this accent modification will be investigated, with a particular focus on the articulation of intervocalic /t/. As a sociolinguistic variable, the realization of this phonetic feature as a flap (closer to the ear to a [d] than a prototypical stop [t]), indexes both an American identity, and an alignment with supra-local pronunciation norms characteristic of popular music, referred to as Pop Song English. This part examines sociocultural causes, including the American dominance of the music industry, the importance of musical inspirations, and the resulting conflicting identities artists may express. An acoustic analysis of other markers of American English (the TRAP and LOT vowels), along with an investigation of how the perception of /t/ has evolved among British speakers situate the study within a broader context. The second component of the thesis explores a complementary articulatory explanation for why British artists may favor a flap /t/ realization: the phonetic constraints that distinguish speech from singing. A novel experimental approach in the study of popular music using several imaging techniques, including lip tracking, electropalatography (EPG), and magnetic resonance imaging (MRI) is implemented with the participation of five professional British singers. These methods are used to examine lip aperture, linguopalatal contact, and overall vocal tract configuration to test the influence of factors such as pitch and phoneme duration on articulation in singing. The underlying hypothesis is that the more extreme these conditions become (at higher registers common in heavy metal, for instance), the less they favor a clear occlusion of the tongue against the alveolar ridge, the more the realization of /t/ tends to shift toward a flap. Thus, the linguistic modification that occurs between speech and singing may also be a strategy of articulatory facilitation in a complex activity in many respects. Results show that British singers' adoption of American-like flapping in song arises from a combination of social-indexical orientation toward the American model and global pop norms, as well as articulatory adjustments specific to singing. They also demonstrate that within this context, individual variation, however subtle at the articulatory level, prevails in the form of personal attitudes and distinct physiological strategies.

Résumé en français

L'ANALYSE SCIENTIFIQUE de la voix chantée est une discipline bien établie (Sundberg, 1987) qui s'attache à employer des méthodes d'imagerie ou d'analyse du signal toujours plus performantes pour expliquer les mécanismes physiologiques impliqués dans les différentes formes de chants, principalement dans le chant lyrique. La sociophonétique, quant à elle, se consacre aux accents, et donc à l'étude de la manière dont l'articulation des sons de la parole révèle l'identité sociale ou culturelle des locuteurs (Labov, 1972; Trudgill, 1986). Si la voix chantée a bel et bien été parfois étudiée dans une optique sociophonétique, à commencer par l'étude novatrice de Trudgill sur l'américanisation de la pop en 1983 (Trudgill, 1983), il peut être délicat d'y associer des méthodes expérimentales et instrumentales de pointe. Et pour cause : ce sont bien souvent des enregistrements d'artistes célèbres qui sont analysés, où la voix a subi divers traitements (compression, effets de réverbération, ajustement de hauteur, etc.) et est mélangée aux autres instruments. Les études sur la voix chantée en sociophonétique ont toujours suivi une approche socio-culturelle, selon laquelle adopter un accent relève d'une démarche délibérée d'identification à une communauté particulière. Beal (2009), Duncan (2017), and Gibson (2019) ont cherché à comprendre ce choix en fonction de groupes (Arctic Monkeys), genres musicaux (la country) ou encore de toute une zone géographique (les chanteurs de musique populaire de Nouvelle-Zélande), respectivement. Ces études et d'autres encore ont permis d'identifier des traits de prononciation saillants, qui mettent en évidence que le fait d'américaniser sa prononciation est bien ancré dans la musique populaire du fait de la domination culturelle des États-Unis dans ce domaine (Trudgill, 1983). On note aussi plus récemment le développement d'une nouvelle approche de ce phénomène. Au lieu d'une américanisation délibérée de la prononciation, les artistes de musique populaire adoptent par réflexe une variété de l'anglais propre au chant, inspirée de l'accent américain, certes, mais qui a perdu toute association directe avec celui-ci. Ce que l'on peut considérer comme une nouvelle variété d'anglais à part entière a été théorisé par Gibson (2023) sous le nom de Pop Song English. Les chanteurs de musique populaire sont contraints de se positionner face à cette norme, en l'adoptant ou la rejetant en revendiquant l'utilisation d'accents régionaux dans le chant.

Malgré la grande variation dans les accents de l'anglais, il est d'usage de prendre comme référence le USA-5 model (Simpson, 1999), qui identifie cinq traits phonétiques qui distinguent l'anglais britannique et l'anglais américain :

- **T-voicing** : le /t/ intervocalique comme dans *better* réalisé en battue se rap-

proche du [d] plutôt que [t] ou [ʔ], plus typiquement britanniques.
- **Absence de TRAP-BATH split**: la voyelle dans les mots comme *ask* et *bath* est réalisée en [æ] plutôt que [ɑː].
- **Rhoticité** : le /r/ non pré-vocalique est prononcé alors qu'il ne l'est pas en anglais britannique.
- **PRICE diphthong** : la diphtongue [aɪ] de *eyes* et *life* est réalisée en [ɑː].
- **LOT unrounding** : la voyelle [ɒ] de *rock* ou *stop* est désarrondie en [ɑː].

Au-delà de la dimension socioculturelle, des contraintes articulatoires influencent également ce changement d'accent entre voix parlée et voix chantée. Morrissey (2008) suggère que certains éléments de prononciation se prêtent mieux au chant, et émet l'hypothèse d'une meilleure « chantabilité » (singability) de certains traits phonétiques au détriment d'autres. Selon la théorie d'Hyper et Hypo-articulation (H&H) de Lindblom (1990), la parole est adaptative : les locuteurs accordent habituellement leur façon de parler en fonction d'exigences communicatives et contextuelles, en variant le long d'un continuum allant de l'hyper à hypo-articulation. Dans la même veine, Lindblom affirme que la parole, en tant qu'elle relève du contrôle moteur humain, a tendance à choisir le comportement qui demande le coût énergétique moindre. En d'autres termes, s'il n'est pas contraint par d'autres facteurs, un locuteur penchera naturellement vers une forte réduction articulatoire et une économie des gestes de la parole. Si cela est vrai pour la voix parlée, on peut facilement imaginer qu'il en est de même pour le chant, qui exige d'autant plus de concentration, énergie et précision que la parole (Sundberg, 1987).

Dans ce contexte, cette thèse cherche à fournir une vision holistique du comportement linguistique des chanteurs de musique populaire britannique et à évaluer dans quelle mesure les approches socioculturelles rendent compte des tendances observées. Une attention particulière est accordée aux années 1980, envisagées comme une période de transition marquant un éloignement de l'influence américaine, tant pour l'industrie musicale en général que sur le plan des modèles linguistiques. Cette analyse repose sur l'étude approfondie d'une sélection d'enregistrements studios issus de deux genres de la musique populaire britannique : le heavy metal traditionnel et la musique pop britannique. Des travaux antérieurs se sont penchés sur ce sujet, mais ils sont souvent d'une portée limitée. La plupart se concentrent sur quelques chansons d'un seul artiste, ce qui conduit parfois à des conclusions générales qui ne rendent pas pleinement compte de la complexité du phénomène étudié. Cette thèse cherche à dépasser cette limite en adoptant une approche plus large, susceptible d'apporter un éclairage supplémentaire sur la sociophonétique de la voix chantée dans la musique populaire.

Un autre objectif de cette approche est d'introduire une perspective supplémentaire dans l'étude de la prononciation en musique populaire : les spécificités de la voix chantée, avec un accent particulier sur les contraintes articulatoires. Cet aspect a jusqu'à présent été largement négligé par la majorité des travaux sur le sujet. Pour combler cette lacune, cette étude adopte une approche expérimentale inédite impliquant la participation de cinq chanteurs de musique populaire britannique. Des techniques telles que le *lip tracking*, l'électropalatographie (EPG) et l'imagerie par résonance magnétique (MRI) sont mises en œuvre afin d'examiner les aspects articulatoires liés à l'ouverture buccale, au contact linguopalatal et à la configuration générale du conduit vocal. L'objectif est d'explorer comment des variables telles que la hauteur, l'intensité

ou la durée des voyelles peuvent influencer le chant et entraîner des différences de prononciation par rapport à la parole. L'hypothèse principale de cette partie expérimentale repose sur l'idée que le chant est un mode de production vocale plus exigeant que la parole, et de ce fait, plus les conditions en sont extrêmes (comme pour un genre musical tel le heavy metal), plus un chanteur tentera de minimiser l'effort vocal requis et aura ainsi tendance à prononcer un /t/ battu, plus facile par ses caractéristiques articulatoires à chanter qu'un /t/ occlusif.

Cette thèse est divisée en trois parties. Dans la Partie I, nous examinons la revue la littérature existante afin de fournir le cadre théorique de ce travail et de le situer dans un contexte de recherche à la fois vaste et pluridisciplinaire. L'analyse de la prononciation du /t/ chez les chanteurs britanniques doit nécessairement s'appuyer sur des recherches issues de domaines tels que la sociolinguistique, la phonétique articulatoire, acoustique et expérimentale, la sociologie, les études sur la musique populaire et la musicologie. Chaque chapitre de cette partie présente un aperçu d'une approche théorique qui relève de l'étude de la voix chantée, sans pour autant prétendre à l'exhaustivité.

Chapitre 1

Le premier chapitre débute avec une introduction générale à la variation sociolinguistique, en s'intéressant plus particulièrement aux concepts de style et de *style-shift* (modification stylistique). Ce chapitre met en évidence l'importance du public et de l'agentivité des locuteurs dans la construction de leurs identités, considérées comme les forces motrices du *style-shift*. Dans le cadre de la troisième vague de la sociolinguistique variationniste, les locuteurs ne sont plus envisagés comme de simples porteurs passifs de traits dialectaux ; ils mobilisent ces traits de manière différenciée, à différents moments, afin de construire certaines identités ou de se positionner par rapport à certains groupes–ou en opposition à ceux-ci. La variation linguistique est profondément ancrée dans la construction du sens social et de l'identité, et les performances de musique populaire offrent un point de vue privilégié pour l'étude de ce phénomène en sociolinguistique.

Chapitre 2

Ce chapitre présente les grandes lignes méthodologiques de l'approche sociophonétique de la voix chantée adoptée dans ce travail, en commençant par un aperçu de l'étude pionnière de Trudgill (1983) et des travaux existants. Puis, une revue de la littérature sur le sujet examine l'adoption ou le rejet d'un style de prononciation influencé par l'accent américain comme outil d'identification, que cette identification soit perçue et/ou projetée. L'évolution théorique allant de « l'américanisation » au Pop Song English suit les transformations historiques de l'industrie musicale dans les années 1980, mises en avant comme un tournant majeur dans la musique populaire et qui seront abordées tout au long de ce travail.

Chapitre 3

L'étude de la voix chantée nécessite d'être appréhendée en relation avec la parole, tout en insistant sur ses caractéristiques distinctives. Les différences entre la prononciation parlée et chantée des chanteurs britanniques peuvent être influencées non seulement par des facteurs socioculturels, mais aussi par des contraintes propres à la voix chantée. Les

notions d'économie vocale et de préférence pour la sonorité (sonority) misent en avant dans ce chapitre suggèrent que certains choix de prononciation peuvent être motivés par le besoin d'optimiser l'effort vocal dans le chant, puisque le système phonatoire est sollicité dans des conditions plus extrêmes que dans la parole ordinaire.

Chapitre 4

L'occlusive /t/ intervocalique à l'intérieur d'un mot, choisie comme variable linguistique pour ce travail, constitue un point d'ancrage idéal : elle est suffisamment spécifique pour permettre une analyse détaillée, tout en étant pertinente dans plusieurs domaines, ce qui en fait un objet utile pour aborder les différents aspects de cette étude. Elle représente à la fois un trait marqué de la prononciation influencée par l'américain et/ou par le PSE, et, par ses variantes, un intérêt sur le plan articulatoire qui offre l'occasion d'examiner comment les contraintes propres au chant peuvent influencer sa réalisation. En effet, le /t/ occlusif nécessite un rétrécissement du conduit vocal, un blocage complet du flux d'air intra-oral grâce à l'écrasement du bout de la langue contre la zone alvéolaire, suivi d'un relâchement soudain de ce blocage. Par opposition, le /t/ battu ne nécessite ni occlusion, ni relâchement du blocage, et la langue ne rentre en contact que brièvement (ou pas du tout) avec la zone alvéolaire.

Dans la Partie II, nous examinons la modification dans la production orale des chanteurs britanniques en mobilisant des perspectives socioculturelles classiques, à travers une étude approfondie d'exemples réels tirés du heavy metal britannique traditionnel et de la pop britannique des années 1980.

Chapitre 5

Ce chapitre propose une analyse quantitative de quatre groupes de heavy metal britannique traditionnel, en se concentrant sur la réalisation du /t/ intervocalique. Les résultats montrent que tous les groupes utilisent la variante battue dans le chant, malgré son absence dans la parole. Les raisons de ce changement de style, motivé par les attentes du public, les normes du genre et des notions d'authenticité, sont considérées. Bien que les tendances diachroniques soient moins concluantes, les résultats mettent en évidence que même un genre culturellement distinct comme le heavy metal reflète des tendances de prononciation plus larges et un alignement potentiel avec les normes émergentes de la musique populaire.

Chapitre 6

En suivant une méthodologie similaire à celle du Chapitre 5, ce chapitre propose une analyse d'un échantillon représentatif de la musique pop britannique des années 1980. Les résultats rejoignent ceux du précédent chapitre en montrant que la réalisation battue du /t/ intervocalique est largement présente dans les productions chantées, mais rare dans la parole, confirmant son rôle de variable linguistique signalant soit unalignement avec les normes inspirées de l'anglais américain, soit une résistance à celles-ci. Les résultats suggèrent que cette décennie marque une transition entre l'adoption explicite d'accents américains et une norme de chant pop plus implicite et standardisée (PSE), les artistes devant négocier entre les pressions du marché, leurs sources d'inspiration musicale et le désir de maintenir une identité britannique distincte.

Chapitre 7

Ce chapitre propose une analyse acoustique exploratoire de la production de certaines voyelles dans le chant, en se concentrant sur les voyelles de TRAP-BATH et LOT dans les performances chantées de quatre artistes britanniques issus de différents genres, et deux chanteurs américains servant de référence. Les résultats montrent que des voyelles dont la réalisation est proche de celles de l'anglais américain apparaissent dans le chant britannique, confirmant que le *style-shift* ne se limite pas aux consonnes comme le /t/ intervocalique. Ce chapitre montre aussi qu'il est possible d'obtenir des mesures acoustiques fiables à partir d'enregistrements préexistants en utilisant des outils de séparation de sources et des méthodes conventionnelles d'extraction des formants.

Chapitre 8

Ce chapitre examine la façon dont les auditeurs britanniques perçoivent l'allophone battu du /t/ intervocalique, en se demandant s'il est encore perçu actuellement comme un trait spécifiquement américain. Les résultats, issus à la fois d'une expérience de perception par identification et d'une enquête, ne révèlent pas de différences de génération majeures : les participants associent de manière constante le /t/ battu à l'anglais américain. Bien que les auditeurs plus jeunes semblent légèrement plus enclins à le percevoir comme un allophone possible de l'anglais britannique, l'étude suggère que la réalisation battue du /t/ reste un marqueur saillant de la parole américaine. Des recherches complémentaires dans un contexte musical permettraient d'enrichir cette perspective.

La Partie III applique des méthodes expérimentales pour tester l'hypothèse articulatoire posée dans la Partie I, selon laquelle certains traits du /t/ battu seraient plus 'faciles' à chanter, a fortiori dans des conditions plus extrêmes. Trois techniques ayant leurs avantages propres sont utilisées : le *lip tracking*, l'électropalatographie (EPG) et l'imagerie par résonance magnétique (IRM). Cinq chanteurs de musique populaire britannique ont participé à ces études de production.

Chapitre 9

Ce chapitre présente la première expérience de production de la thèse, visant à confirmer le rôle de l'ouverture labiale dans le chant par rapport à la parole, ainsi que l'influence de l'intensité et de la hauteur. Les résultats confirment en grande partie les hypothèses testées, en montrant des différences systématiques d'ouverture labiale entre la parole et le chant, et en révélant que la hauteur et l'intensité du chant influencent de manière significative l'ouverture des lèvres, fournissant ainsi une base pour les expériences ultérieures.

Chapitre 10

Ce chapitre présente l'expérience centrale de la thèse, qui utilise l'électropalatographie (EPG) pour étudier comment les contraintes articulatoires du chant influencent la réalisation du /t/ intervocalique. L'étude compare la parole et le chant, teste les effets de la durée des syllabe et de la hauteur, et montre que, si certaines tendances confirment les hypothèses (les syllabes plus courtes et les hauteurs plus élevées tendent à réduire le

contact linguopalatal), les résultats restent inconstants et fortement influencés par des stratégies articulatoires individuelles. Des analyses détaillées par participant révèlent une variation inter-locuteurs importante : certains chanteurs présentent une réalisation en battue du /t/ et un contact réduit, tandis que d'autres conservent une production stable et prototypique du /t/.

Chapitre 11

Ce chapitre présente une étude exploratoire utilisant l'IRM pour observer la configuration du conduit vocal et la dynamique de la mâchoire pendant la parole et le chant. Les données de deux chanteurs montrent une ouverture mandibulaire généralement plus grande en chant qu'en parole, ce qui correspond aux résultats du Chapitre 9. Peu de /t/ battus ont été produits, et la résolution temporelle limitée du système a mené à une analyse plus restreinte que prévu, ce qui peut expliquer en partie pourquoi aucune autre différence marquante de configuration du conduit vocal entre le chant et la parole n'a été trouvée.

Cette thèse propose un compte rendu socioculturel et articulatoire des contraintes qui régissent la prononciation dans le chant de la musique populaire. Á travers l'utilisation de techniques traditionnelles d'analyse de corpus et une approche expérimentale novatrice dans le domaine, elle donne un exposé détaillé de la sociophonétique de la voix chantée en musique populaire et du rôle des choix linguistiques et contraintes articulatoires dans la construction de l'identité d'un groupe de musique ou d'un chanteur. En d'autres termes, les forces socioculturelles, associées aux contraintes physiques et articulatoires, interagissent dans un processus dynamique qui façonne la prononciation de la voix chantée.

Acknowledgments

I would like to express my sincere gratitude to Emmanuel Ferragne for your guidance and mentorship throughout these four (seven, really!) years. I came to you a brand new grad student determined not to do a master's in phonetics... How wrong I was. Thank you for seeing the potential in me and introducing me to the tough but gratifying world of research, and specifically experimental methods. I've accomplished so many things during this PhD I had never even dreamed I could do. Thank you for your trust, encouragement, and unfaltering availability. I don't know if you noticed anything special with the chapters in my thesis but, believe it or not, they go up to 11.

I wish to thank Joan Beal, Claire Pillot-Loiseau, Jane Stuart-Smith, and Sylvain Navarro for giving me the honor of accepting to be members of the jury to examine this work. Many thanks to Claire and Sylvain, as members of my committee, for your precious input and support throughout this process.

A warm thank you to the other members of the Université Paris Cité karaoke linguists crew, Hannah King, Manon Lelandais, Maud Pélissier, and Anne Talbot for being both inspiring role models and great singing partners during our weekly lunch jamming sessions. Many thanks also to Corinne Hamel for helping me navigate the administrative side of this work.

There are so many people without whom none of this would have been possible. I want to express my gratitude to Paul Sharp from icSpeech for his precious help with the EPG, as well as the members of Laboratoire Fraisse, Cécile, Nathalie and Nadia for the pseudopalate creation. To Clément Debacker, Maliesse Lui, and especially Sylvain Charron, thank you for your help with all things MRI. To Eric Ferrari, your help was greatly appreciated and I hope this gave you a positive first experience of research.

A very special thanks to Duncan, Josh, Nic, Pat, and Tim for your curiosity and enthusiasm for this project!

People often say that a PhD is a long and solitary process. I can vouch for the length of it, but definitely not for the loneliness part! Throughout these four years I've had the chance to be surrounded in the PhD student's office by the best colleagues I could have asked for. I'm so grateful for the laughs, the tears, and everything in

between I've shared with my room 825 family. Thank you to Léa, Léo, Alessandro, and Cameron for making the first half of my journey so smooth. And of course, thank you to those who were there throughout the (bumpier) second half: Alice, Behnoosh, and Tacita. Alice, I'll catch you on the flip side, wherever that may be! To Leela, too, it was comforting to know there was someone out there going through the same thing at the same time.

Thank you to all my friends for their unwavering support and understanding when I disappeared to finish writing for a few months, I can't name all of you here but you're all in my heart.

To my mom and dad, for being my safe space whenever I need a break from everything, and always being there for me no matter what. To Auguste and to new beginnings. To Céleste and to us both becoming (very different kinds of) doctors. To Pierre, for starting my musical education, and to Archi, for perfecting it. To my entire family. You mean the world.

To Hélène Devissaguet and Jean-Michel Massou dit Bourdet. You have, both in your own ways, led me to exactly where I am today, and for that I am beyond grateful.

And to Tom, thank you for being by my side throughout, agreeing to be my test subject when I needed one, and being so involved in the final process. You've probably done the hardest part of all with your edits, and I can never blame you again for not reading my work. This is (almost) your *thèse* as much as mine.

Preface

In order to improve the reader's experience of this thesis, a few things should be noted:

Definitions of important terms are provided in a Glossary, located just before the General Introduction. A List of Acronyms is provided in the same section. In addition, a partial Index of relevant terms and their corresponding page numbers appears at the end of the manuscript, following the Bibliography.

The electronic version of this work includes hyperlinks, indicated when words appear in dark red. Depending on the nature of the link, they will direct the reader to the Glossary entry, the List of Acronyms, or the referenced chapter or section. After following a link, readers can return to their previous location in the document by simultaneously pressing the `alt` and `left arrow` keys on their keyboards in Windows and Linux, or the `command` and `[` keys on macOS. Bibliographic references are clickable as well, on the year of publishing. Links to external websites are indicated in pink.

Table of Contents

Table of Contents **xxiii**

List of figures **xxvii**

List of tables **xxxi**

List of acronyms **xxxiii**

Glossary **xxxv**

General introduction **1**

I Background **5**

1 Sociolinguistic Variation **7**

1.1 The Three Waves of sociolinguistic variation 8

1.2 Style 10

1.2.1 Style-shift 11

1.2.2 Sociolinguistic theories of style 12

1.3 Sociolinguistic variation and pop culture 17

1.3.1 Pop culture: an object of sociolinguistic study 17

1.3.2 Sociolinguistics and popular music 19

1.4 Summary and conclusion 22

2 The Sociophonetics of popular music singing **23**

2.1 Style-shift in popular music singing 24

2.1.1 An overview of Trudgill's *Sociolinguistics of British Pop-song Pronunciation* 24

2.1.2 After Trudgill . . . 26
2.2 Theoretical implications . . . 30
2.2.1 Sociophonetic processes . . . 30
2.2.2 Theories of style applied to singing . . . 31
2.3 Style-shifting in context . . . 33
2.3.1 From 'Americanization' to Pop Song English . . . 33
2.3.2 A chronology of singing pronunciation style in popular music . . . 35
2.4 Summary and conclusion . . . 37

3 An Articulatory Perspective 39
3.1 Articulatory framework . . . 40
3.1.1 H&H Theory . . . 40
3.1.2 Economy of speech gestures . . . 41
3.2 The specificity of the singing voice . . . 42
3.2.1 Singing versus speech . . . 42
3.2.2 Styles of singing . . . 44
3.3 Articulatory constraints applied to singing . . . 48
3.3.1 Vocal economy in singing . . . 48
3.3.2 A preference for sonority . . . 50
3.4 Summary and conclusion . . . 52

4 The Case of Intervocalic /t/ 53
4.1 The pronunciation of intervocalic /t/ in English . . . 54
4.1.1 Allophones of intervocalic /t/ . . . 54
4.1.2 Flapping in focus . . . 61
4.2 Sociophonetic aspects of intervocalic /t/ . . . 62
4.2.1 Flapping: a defining feature of American English? . . . 63
4.3 Intervocalic /t/ in singing . . . 65
4.3.1 Features of stop /t/ vs. flap /t/ in singing . . . 66
4.4 Summary and conclusion . . . 68

Part I: Summary and Research Questions 69
4.5 Summary . . . 69
4.6 Research questions . . . 70

II Sociocultural factors 73

Part II: General Introduction 75

5 Heavy Metal Pronunciation 77
5.1 Background . . . 78
5.1.1 Why study traditional British heavy metal? . . . 78
5.1.2 Selected artists . . . 79
5.1.3 Aims . . . 80
5.2 Methodology . . . 81
5.2.1 Corpus . . . 81
5.3 Analysis . . . 82
5.3.1 Pre-processing . . . 82
5.3.2 Statistical analysis . . . 83
5.4 Results and discussion . . . 84
5.4.1 Results . . . 84
5.4.2 Discussion . . . 86
5.5 Summary and conclusion . . . 94

6 Pop Music Pronunciation 95
6.1 Background . . . 96
6.1.1 Pop music . . . 96
6.2 Aims . . . 96
6.3 Methodology . . . 96
6.4 Analysis . . . 98
6.4.1 Pre-processing . . . 98
6.5 Results and discussion . . . 100
6.5.1 Results . . . 100
6.5.2 Discussion . . . 103
6.6 Summary and conclusion . . . 107

7 A Complementary Acoustic Analysis 109
7.1 Background . . . 110
7.1.1 Acoustic analysis of the singing voice . . . 110
7.1.2 Selected artists . . . 110
7.1.3 Selected vowels . . . 111
7.1.4 Aims and predictions . . . 112
7.2 Methodology . . . 113
7.2.1 Corpus . . . 113
7.3 Analysis . . . 115
7.3.1 Pre-processing . . . 115
7.3.2 Acoustic analysis . . . 115

7.3.3 Reliability analysis . . . 115
7.3.4 Statistical analysis . . . 116
7.4 Results and discussion . . . 117
7.4.1 TRAP-BATH . . . 117
7.4.2 LOT . . . 119
7.4.3 Classification accuracy: formants vs. MFCC . . . 119
7.4.4 Discussion . . . 120
7.5 Summary and conclusion . . . 122

8 The perception of intervocalic /t/ 123
8.1 Background . . . 124
8.1.1 Perception of singing voice pronunciation . . . 124
8.1.2 Aims and predictions . . . 124
8.2 Methodology . . . 125
8.2.1 Participants . . . 125
8.2.2 Stimuli . . . 126
8.2.3 Procedure . . . 127
8.3 Analysis . . . 129
8.3.1 Preprocessing . . . 129
8.3.2 Statistical analysis . . . 130
8.4 Results and discussion . . . 131
8.4.1 Identification accuracy . . . 131
8.4.2 Response time . . . 132
8.4.3 Survey . . . 133
8.5 Discussion . . . 135
8.6 Summary and conclusion . . . 138

Part II: Summary 141

III Articulatory factors 143

Part III: General Introduction 145

9 Production Experiment 1: lip tracking 147
9.1 Background . . . 148
9.1.1 Mouth and lip aperture in singing . . . 148
9.1.2 Aims and predictions . . . 148
9.2 Methodology . . . 149

9.2.1 Methods for lip tracking . . . 149
9.2.2 Participants . . . 152
9.2.3 Stimuli . . . 154
9.2.4 Experimental setup and material . . . 157
9.2.5 Procedure . . . 159
9.3 Analysis . . . 160
9.3.1 Preprocessing . . . 160
9.3.2 Extraction . . . 161
9.3.3 Statistical analysis . . . 162
9.4 Results and Discussion . . . 163
9.4.1 Speaking versus singing comparison . . . 163
9.4.2 Singing only analysis . . . 164
9.4.3 Discussion . . . 167
9.5 Chapter summary and conclusion . . . 168

10 Production Experiment 2: EPG 169
10.1 Background . . . 170
10.1.1 What is electropalatography? . . . 170
10.1.2 The LinguaGraph unit . . . 172
10.1.3 Electropalatographic studies . . . 173
10.1.4 Aims and predictions . . . 178
10.2 Methodology . . . 178
10.2.1 Participants . . . 178
10.2.2 Stimuli . . . 179
10.2.3 Material and experimental setup . . . 182
10.2.4 Procedure . . . 183
10.3 Analysis . . . 185
10.3.1 Preprocessing . . . 185
10.3.2 EPG data extraction and reduction . . . 187
10.3.3 Pseudopalate normalization . . . 190
10.3.4 Relabeling pitch . . . 192
10.3.5 Statistical analysis: a caveat . . . 193
10.4 Results and Discussion . . . 194
10.4.1 Eliciting flaps: results . . . 195
10.4.2 Eliciting flaps: discussion . . . 199
10.4.3 Speaking vs. singing comparison: results . . . 201
10.4.4 Speaking vs. singing comparison: discussion . . . 203
10.4.5 Linguopalatal contact: results . . . 205

10.4.6 Linguopalatal contact: discussion 214
10.4.7 A closer look at some tokens 216
10.5 Summary and Conclusion 219

11 Production Experiment 3: MRI **223**
11.1 Background 224
11.1.1 Magnetic Resonance Imaging (MRI) 224
11.1.2 Aims and predictions 225
11.2 Methodology 226
11.2.1 Participants 226
11.2.2 Stimuli 226
11.2.3 Equipment 227
11.2.4 Procedure 227
11.3 Analysis 229
11.3.1 Data preprocessing 229
11.3.2 Articulatory alignment 231
11.3.3 Distance estimation 232
11.4 Results and discussion 233
11.4.1 Eliciting flaps 234
11.4.2 Speaking vs. singing comparison: results 234
11.4.3 Speaking vs. singing comparison: discussion 238
11.4.4 A closer look at one occurrence of flapping 241
11.5 Chapter summary and conclusion 242

Part III: Summary **245**

General conclusion **247**
11.6 Main findings 247
11.7 Contributions & Implications 251
11.8 Limitations and future directions 252
11.9 Conclusion 255

Appendices **259**

A Part II Appendix **259**
A.1 Chapter 5: Heavy metal pronunciation 259
A.1.1 Albums included in the corpus 259
A.1.2 Albums by American bands included in the corpus 260

A.1.3 Interviews . . . 261
A.2 Chapter 6: Pop music pronunciation . . . 264
A.2.1 Albums, selected songs, and interviews included in the corpus . 264
A.3 Chapter 7: Acoustic analysis . . . 272
A.3.1 Albums included in the corpus . . . 272

B Production Experiment 1: lip tracking 275
B.1 Ethics Committee Approval . . . 275
B.2 lip tracking by color segmentation MATLAB script . . . 277
B.3 lip tracking by facial landmark recognition Python script . . . 279
B.4 Participant information form . . . 281
B.5 Recruitment flyer . . . 284
B.6 Participant information letter . . . 286
B.7 Melody scores . . . 291
B.8 Experiment stimuli display MATLAB script . . . 293
B.9 Experiment setups for lip tracking . . . 304
B.10 Experiment consent form . . . 305

C Production Experiment 2: EPG 309
C.1 Questions for semi-guided interview . . . 309
C.2 Abbreviated transcripts of participant interviews . . . 311
C.2.1 Participant s01 . . . 311
C.2.2 Participant s02 . . . 313
C.2.3 Participant s03 . . . 314
C.2.4 Participant s04 . . . 316
C.3 EPG data extraction and reduction MATLAB script . . . 320
C.4 Models . . . 323

D Production Experiment 3: MRI 329
D.1 Recruitment flyer . . . 329
D.2 MRI DICOM to video MATLAB script . . . 331
D.3 Image alignment MATLAB script . . . 334
D.4 Distance measures from image MATLAB script . . . 337

Bibliography 341

Index 361

List of Figures

1.1 Style as audience and referee design 14

3.1 An illustration of the H&H Theory 41
3.2 The three functional constituents of the voice organ 43
3.3 Proportions of phonation time in speech and singing 44
3.4 The sonority scale 51

4.1 Prototypical stop 55
4.2 Lingual overshoot in the production of [t] 56
4.3 Prototypical flap 58
4.4 Diagrams of the four flap/tap variants 60
4.5 Electroglottography waveforms of stop /t/ and flap /t/ in the word *little* 67

5.1 Proportion of /t/ realizations in heavy metal corpus 84
5.2 Boxplot of token duration according to /t/ realization. 86
5.3 Proportion of flap /t/ realizations per studio album per band 87

6.1 Percentage of flapped intervocalic /t/ for each artist in the pop corpus. 101
6.2 Token duration (ms) according to /t/ realization 102

7.1 Mixed track vs. isolated vocal track comparison 114
7.2 cp_formants GUI 116
7.3 F1 and F2 values for TRAP, BATH and LOT vowels 118

8.1 Example trial from perception experiment 128
8.2 Proportion of perception match 131
8.3 Answers to survey question 4 134

9.1 Lip tracking using automatic color segmentation 150

9.2 Example of unsuccessful lip tracking segmentation outputs 151
9.3 Example of outputs from facial landmark detection 152
9.4 Experimental flow 158
9.5 Violin plot of lip aperture according to voice condition 164
9.6 Predicted lip aperture heatmap 165
9.7 Individual intensity, pitch, and lip aperture trends per participant . . . 166

10.1 Diagram of an electropalatography system 170
10.2 Example of static palatography 171
10.3 LinguaGraph pseudopalate 174
10.4 EPG tongue-palate contact pattern for a prototypical alveolar stop /t/ 177
10.5 Diagram of stimuli conditions for Production Experiment 2 181
10.6 Mains steps of the EPG palate creation process 184
10.7 Examples of EPG contact visualization 189
10.8 Result of EPG data refining functions 190
10.9 Pitch clusters according to participant 193
10.10 Proportion of /t/ realizations in sung tokens 195
10.11 Flapping proportion according to pitch and syllable duration 196
10.12 Predicted flap probabilities across conditions 198
10.13 Whole palate contact according to voice condition 202
10.14 Anterior contact according to voice condition 204
10.15 Diagram of linguopalatal contact according to experimental conditions 207
10.16 Anterior linguopalatal contact according to pitch 208
10.17 Normalized anterior contact according to pitch per participant 208
10.18 Normalized anterior contact for three syllable duration 209
10.19 Normalized anterior contact by syllable duration per participant 210
10.20 Whole palate activation for stops and flaps for s02 211
10.21 Variance heatmaps per participant 212
10.22 Average linguopalatal contact trajectories 213

11.1 MRI experimental setup 228
11.2 MRI experimental flow 230
11.3 Pixel selection in Image Viewer App in MATLAB 233
11.4 Overlay of sung vs. spoken MRI frames for s04 235
11.5 Overlay of sung vs. spoken MRI frames for s05 236
11.6 Z-score normalized distribution of jaw opening proxy distances 237
11.7 Jaw opening distances by word and voice condition 240
11.8 Jaw opening by vowel quality and voice condition per participant . . . 241

11.9 Overlay of flap /t/ in singing and speaking 242
11.10Relative contribution of articulatory and sociocultural factors 251

List of Tables

2.1 The USA-5 model 24

4.1 Four flap and tap variants in English 59
4.2 /t/ realizations by British English speakers in two phonological contexts 65

5.1 Albums, songs, and interviews included in the heavy metal corpus . . . 81
5.2 Distribution of flaps and stops in the heavy metal corpus 85
5.3 Output of a logistic regression model predicting flapping as a function of log-transformed token duration and band 88

6.1 Subset of artists and albums included in the pop song corpus 98
6.2 Output of a generalized linear mixed model predicting flapping as a function of log-transformed token duration. 102

7.1 Album and song counts for artists in the acoustic analysis corpus. . . . 113
7.2 Average Euclidean distance between LOT and TRAP vowels in F1-F2 . . 119
7.3 Classification rates for differentiating TRAP-BATH from LOT vowels in percentages, using SVM 120

8.1 72 nonce words created as stimuli for Perception Experiment 1, inspired by Braver (2014). 127
8.2 Output of a generalized linear mixed-effects model predicting match of flap realization to orthographic <t> as a function of respondent Generation and Accent of the audio stimuli. 132
8.3 Output of a linear mixed-effects regression model predicting response time as a function of age group, accent condition, and accuracy of response and their interaction. 133
8.4 Notable American features according to British respondents 135

9.1 Inclusion criteria for participants in production experiment 153
9.2 Stimuli for Experiment 1 156
9.3 Number of tokens available for analysis per participant for Experiment 1. 161
9.4 Output of a linear mixed-effects regression model predicting lip aperture (z-scored) as a function of voice condition. 163
9.5 Output of a linear mixed-effects regression model predicting lip aperture as a function of pitch and intensity. 164

10.1 Overview of EPG studies 175
10.2 Duration (ms) of /t/ in flaps and stops for heavy metal and pop 180
10.3 Number of tokens available for analysis per participant for Experiment 2. 187
10.4 Mean alveolar activation from pseudopalate calibration tests. 191
10.5 Whole-recording activation percentages by participant and region . . . 192
10.6 Summary of variables used in statistical modeling. 194
10.7 Top five most flapped words 196
10.8 Output of a generalized linear model predicting flapping as a function of syllable duration and pitch for participant s02 197
10.9 Output of a linear mixed-effects regression model predicting linguopalatal contact for the whole palate as a function of voice condition 201
10.10 Output of a linear mixed-effects regression model predicting linguopalatal contact for the whole palate as a function of voice condition (speaking or singing) for /t/ tokens specifically. 203
10.11 Output of a linear mixed-effects regression model predicting linguopalatal contact for the anterior region as a function of voice condition (speaking or singing) for /t/ tokens specifically. 203
10.13 Variance measures per participant 212
10.14 Tokens according to anterior contact (normalized) and flap realization . 217

11.1 Number of tokens available for analysis per participant for Production Experiment 3. 231
11.2 Average distance between labial tubercle and chin tip 237
11.3 Token groups by vowel quality 239

C.1 Output of a linear mixed-effects regression model predicting linguopalatal contact for the whole palate as a function of pitch and syllable duration. 323
C.2 Output of a linear mixed-effects regression model predicting linguopalatal contact for alveolar and lateral regions as a function of pitch and syllable duration. 323

C.3 Output of a linear mixed-effects regression model predicting linguopalatal contact for the anterior region as a function of pitch and syllable duration. 324
C.4 Output of a linear mixed-effects regression model predicting linguopalatal contact for the whole palate (normalized per participant) as a function of pitch and syllable duration. 324
C.5 Output of a linear mixed-effects regression model predicting linguopalatal contact for alveolar and lateral regions (normalized per participant) as a function of pitch and syllable duration. 324
C.6 Output of a linear mixed-effects regression model predicting linguopalatal contact for the anterior region (normalized per participant) as a function of pitch and syllable duration. 325
C.7 Output of a linear mixed-effects regression model predicting linguopalatal contact for the whole palate as a function of k-means determined relabeled pitch and syllable duration. 325
C.8 Output of a linear mixed-effects regression model predicting linguopalatal contact for alveolar and lateral regions as a function of k-means determined relabeled pitch and syllable duration. 325
C.9 Output of a linear mixed-effects regression model predicting linguopalatal contact for the anterior region as a function of k-means determined relabeled pitch and syllable duration. 326
C.10 Output of a linear mixed-effects regression model predicting linguopalatal contact for the whole palate (normalized per participant) as a function of k-means determined relabeled pitch and syllable duration. 326
C.11 Output of a linear mixed-effects regression model predicting linguopalatal contact for alveolar and lateral regions (normalized per participant) as a function of k-means determined relabeled pitch and syllable duration. 326
C.12 Output of a linear mixed-effects regression model predicting linguopalatal contact for the anterior region (normalized per participant) as a function of k-means determined relabeled pitch and syllable duration. 327

List of acronyms

AmE American English. 51, 62–65, 105, 112, 120, 126, 131, 135, 138

BNC British National Corpus. 218

BrE British English. 51, 62, 64, 65, 112, 119, 126, 130–132, 135, 136, 138, 142, 234, 252

CCM contemporary commercial music. 46–48, 52, 148, 225

COCA Contemporary Corpus of American English. 218

DAW Digital Audio Workstation. 155

EGG electroglottography. 66, 67

EMA electromagnetic articulography. 61

EPG electropalatography. vi, ix, xxvi, 2, 4, 55, 145, 147, 162, 169–174, 176–179, 183–185, 187, 233, 234

fps frames per second. 149, 160, 162

GUI Graphical User Interface. 115

H&H 'Hyper-' and 'Hypo-' articulation Theory. vi, 40, 41, 48, 52, 62

MRI Magnetic Resonance Imaging. 145, 146, 242

NWOBHM New Wave of British Heavy Metal. 89

PC pop culture. 17, 18, 23, 124

PSE Pop Song English. viii, 35, 37, 68, 70, 77, 81, 91, 92, 96, 108, 124, 137, 138, 141, 221

RP Received Pronunciation. 63

RT response time. 130, 132, 136

SAT Speech Accommodation Theory. 12, 13, 22, 31, 33

SSBE Standard Southern British English. 63, 64, 126

SSRH Smooth Signal Redundancy Hypothesis. 40, 41

Glossary

'Hyper-' and 'Hypo-' articulation Theory Theory put forward by Lindblom (1990), which states that speakers are influenced by competing demands, placing them on a spectrum tending towards either listener-oriented hyperarticulated speech or speaker-oriented hypoarticulated speech. 40, 137, 254

Acts of Identity Theorized by Le Page and Tabouret-Keller (1985). Acts of Identity are speech acts used by speakers to externalize different representations of their identity at different times using stylistic variation. Acts of Conflicting Identity refers to the specific adoption of this theory by Trudgill (1983), applied to British popular music singers who must navigate the strength of various and possibly conflicting motivations towards one group or another, including the desire to retain their own unique identity. 15, 17, 22, 25, 31–33, 92, 95, 241

American English In this work specifically, American English (AmE) is similar to Wells (1982a)'s General American variety as a generic and unmarked North American accent. 62–65

apical A term for speech sounds made with the tip of the tongue. 238

approximant A consonant produced by bringing one articulator (the tongue or lips) close to another without actually touching it. 62

Attention to Speech Model Theorized by Labov (1972), the model proposes that the style a speaker uses varies with the level of attention they give to their own speech, which is in turn influenced by the formality of the context. 11

Audience Design Theorized by Bell (1984), a paradigm that posits style-shift occurs primarily in response to the speaker's addressees. 13, 15, 22, 27, 32, 33, 88, 95

belting A powerful singing technique that uses a strong, speech-like quality in the higher part of the voice to produce a loud, resonant sound. 148

British English In this work specifically, British English (BrE) refers to a general variety of English as it is spoken in England. This comprises regional accents but excludes varieties spoken in Ireland, Scotland and Wales. 62–65

Contemporary commerical music Umbrella term referring to modern, non-classical styles of vocal music that are commercially popular. 46

coronal anatomical plane that divides the body longitudinally into dorsal and ventral sections. 172

DICOM Acronym that stands for Digital Imaging and Communications in Medicine. It is a file format containing a single MRI slice and any embedded metadata. 229, 232

Digital Audio Workstation or DAW, a software used to record, edit, arrange, and mix audio on a computer, serving as the main tool for music and sound production. 155

electropalatography A technique for studying articulation that tracks tongue contact with the roof of the mouth using a special artificial pseudopalate with electrodes. 55, 147, 150, 157, 159, 169, 170, 172, 173, 177, 178, 183, 186, 212, 245, 251–253

enregisterment Process by which a way of speaking becomes socially recognized as a distinct variety, style, or register associated with particular groups, contexts, or identities (Agha, 2005). 10, 22, 27

falsetto A vocal register that produces a light, high-pitched sound, typically higher than a singer's normal (modal) voice. 148

First British Invasion A cultural influx in mid-1960s influx of British rock and pop acts into the American music market, catalyzed by The Beatles' huge success. 25, 35, 36, 88, 95, 96

flapping In articulatory terms, flapping refers to a voiced type of articulation where the tongue tip hits the alveolar ridge (or other articulator) very briefly in passing, and continues its movement along a single parallel trajectory. In this thesis, flapping is the umbrella term used to refer to any phonetic realization where /t/

is realized with little to no occlusion and reduced linguopalatal contact, perceived to the ear as closer to a [d] than a prototypical stop [t] . 57–59, 61, 64, 65, 67, 82, 85, 86, 90–92, 100, 102, 106, 107, 123, 125, 135, 138, 141, 142, 196, 197, 199, 200, 216, 220, 225, 245, 248, 250, 252, 254

hypercorrection First coined by Jespersen (1922), occurs when speakers overapply a rule that is perceived as more "correct" in environments where it is actually inappropriate. 16, 31, 32, 54, 105

indexicality The way language or speech features signal social meanings, identities, or contexts. 8, 10, 12, 22, 27, 53

isogloss A geographic boundary that separates areas according to differences in linguistic features. 120

labial tubercle Fleshy protuberance located in the center of the upper lip, directly under the central depression (Gray & Lewis, 1918). 232, 237, 239

laminal A term for speech sounds made with the blade of the tongue (the flat part just behind the tip). 238

lateral tongue bracing Intentional tongue posture whereby the sides of the tongue make contact with the sides of the palate and the upper molars. 54, 55, 66, 176, 187, 213

mandible Lower jaw. It is the largest and strongest bone of the face and serves for the reception of the lower teeth (Gray & Lewis, 1918). 232

MFCC Acronym for Mel-Frequency Cepstral Coefficients. Acoustic features that describe the spectral shape of a sound on a perceptual (Mel) scale. Another way of studying acoustics which does not require formant estimation. 116, 122

MIDI An acronym of Musical Instrument Digital Interface. It is a file format that encodes musical notes and other performance related data but not actual audio recording. 155, 157

mode of phonation Physiological and acoustic setting of the voice when producing sound. Common modes of phonation include modal voice, breathy voice, or falsetto. 43

motor skill An ability involving the movement and coordination of the muscles in the body to perform specific actions. 40

MRI Magnetic Resonance Imaging. A medical imaging technique that uses magnetic fields and radio waves to create detailed pictures of the inside of the body. vi, 2, 4, 223–225, 228, 242, 246, 252–254

New Wave of British Heavy Metal (NWOBHM) Term used to describe the emergence of British heavy metal bands in the mid-1970s who started achieving international fame at the beginning of the 1980s. Examples include Iron Maiden and Def Leppard. 36, 79, 80

occlusal anatomical plane that refers to a top-down view of the teeth and palate, as seen when looking into the mouth from above. 170, 176, 177

overshoot In **sociophonetics**, overshoot refers to when speakers overapply a rule that is perceived as more 'correct' in environments where it is actually inappropriate. In **articulatory phonetics**, it refers to a phenomenon where the articulatory gesture exceeds the target position necessary to produce a particular speech sound. 31, 32, 54, 58, 66, 104, 200

Pop Song English Theorized by Gibson (2023), a supralocal variety of English used in singing, recognized as a standard for most genres of popular music worldwide. v, vii, 35, 69, 92, 96, 106, 123, 141, 248

registration A central technique in medical imaging, registration is the process of establishing spatial correspondences between images. 231, 238, 242, 254

regular expression A formal pattern that defines and locates a specific set of strings in a file. 82, 99, 113, 115

sagittal anatomical plane that divides the body into left and right portions. *Mid*sagittal refers to a section or plane taken exactly through the midline. 49, 54, 59, 61, 172, 227

Second British Invasion The early-to-mid 1980s surge of British pop and new wave artists who gained major popularity in the United States, driven largely by the rise of MTV. 35–37, 79, 88, 96, 121

singability Concept by Morrissey (2008) whereby some speech sounds lends themselves to singing better than others. vi, 39, 50–52, 65

Smooth Signal Redundancy Hypothesis Theorized by Aylett and Turk (2004), the SSRH posits that the constraint of producing robust communication while efficiently expending articulatory effort leads to an inverse relationship between acoustic redundancy and language redundancy. 40, 254

sociolinguistic variable A linguistic feature that can vary across speakers or contexts and thereby index social or stylistic meaning. 8, 25, 53, 68, 77, 81, 141

sonority Refers to the relative loudness or amplitude of a sound, based on its carrying power, acoustic energy, and the degree of freedom of the passage of air through the vocal tract. viii, 44, 50–52, 252

style Intra-speaker variation whereby individuals adopt different linguistic features according to context as a tool for identity-construction. Style-shift refers to the dynamic dimension of this concept, which can change on a moment-to-moment basis. 10, 11, 15, 17, 22, 23, 25, 26, 29–33, 37, 78, 79, 92, 107, 199

support vector machine A machine learning method that separates data into categories by finding the best dividing line between them. 119

tapping In articulatory terms, tapping refers to a voiced type of articulation where the tongue makes one complete closure with the alveolar ridge (or other articulator) and ends its movement in the same relative position. In this work, this type of /t/ realization is encompassed within the term flapping. 57–59

Three Waves of Variation Theorized by Eckert (2012), the Three Waves define the three consecutive approaches that sociolinguistic variation studies have taken over the years. In the first and second wave, variation is seen as marking social categories. In the third wave, variation does not simply reflect, but also constructs social meaning. 8, 15, 22

tomographic Refers to to the creation of cross-sectional or 3D images from multiple projections, as in CT or MRI scans. 223

Top 40 Radio format based on playing repetitively and in well-coordinated rotation the most popular hits of the day. 37

USA-5 model First developed by Trudgill (1983) and later worked into a proper model by Simpson (1999), a list of five defining phonetic variables that differ between British and American English and used by British popular music singers when they style-shift. v, 24, 26, 31, 32, 51–53, 63, 77, 109, 111, 122, 124

vermilion border As defined in J. C. Carey et al. (2009), the vermillion border of the lips corresponds to the rim of paler skin that demarcates the vermilion (the red part of the lips) from the surrounding skin. 152

General introduction

[teɪk ə sæd sɔːŋ ən meɪk ɪt beɾə]

"Hey Jude," The Beatles (1968)

With eight million copies of "Hey Jude" sold worldwide, these lines are some of the most well-known lyrics in popular music. Beyond the words themselves, it is the way Paul McCartney pronounces them that deserves attention. In particular, the realization of the word *better* is striking, since, as a British person, Paul McCartney should most likely pronounce it [betə], with a voiceless alveolar plosive [t], as opposed to a flap [ɾ]. A mere detail to some, this phenomenon captured sociolinguist Peter Trudgill's interest, and led him to conclude that "there can be no doubt that [British] singers are modifying their linguistic behaviour for the purposes of singing" (Trudgill, 1983, p. 143). Besides simply changing their pronunciation from their speaking to their singing voice, it appears that British singers–at least at the time of publication of Trudgill's study–tended toward an American pronunciation model, the realization of /t/ being just one example among others. Any listener of popular music may have identified this tendency, but the reason behind it remains a subject of inquiry for many in the field of sociophonetics and beyond.

Accents can be seen as vectors of identity and tools for creating and maintaining feelings of belonging and differentiation. Accents in the singing voice are no exception. From a traditional sociophonetic approach, British singers adopting American pronunciation features when singing (which they would not use in speech) has been interpreted as a phenomenon that started as a conscious desire to "sound American," and has evolved into a pronunciation standard that most popular music singers default to without active consideration. But the act of singing is also a complex one, differing from speech in terms of vocal effort, aesthetic considerations, and phonatory constraints, among other aspects. Studying singing pronunciation without considering

its inherent production constraints can only provide a partial explanation for phenomena such as McCartney's flap.

The aim of this thesis is twofold and stems directly from this observation. First, it seeks to provide a comprehensive overview of British popular music singers' linguistic behavior, and assess the extent to which sociocultural approaches account for the trends observed. Particular emphasis is given to the 1980s as a period of transition away from American influence, both for the music industry in general, as well as in terms of linguistic models. This is done through an in-depth analysis of a selection of studio productions from two British popular music genres: traditional British heavy metal and British pop music. Previous studies have explored this topic, but they are often limited in scale. Most focus on a few songs by a single artist, sometimes leading to broad conclusions that may not fully capture the complexity of the phenomenon studied. This thesis seeks to overcome that limitation by adopting a more expansive approach, which may provide additional insight into the sociophonetics of the singing voice in popular music. Another motivation for this in-depth approach is to introduce an additional perspective to the study of popular music pronunciation: the specificities of the singing voice, with a particular focus on articulatory constraints. This has so far been largely overlooked by most studies on the subject. To address this gap, this study adopts a novel experimental approach with the participation of five British popular music singers. Techniques including lip tracking, electropalatography (EPG) and MRI are implemented in order to examine the articulatory aspects of lip aperture, linguopalatal contact, and vocal tract configuration, exploring how variables such as pitch, intensity, or vowel duration may impact singing and lead to characteristic pronunciation differences when compared to speech.

Across the corpus-based and experimental research carried out in this thesis, the realization of word-internal intervocalic /t/ is investigated as the main indicator of linguistic modification. The difference between the flapped allophone as heard in "Hey Jude" and its expected stop counterpart is one of the most striking differences between American and British English. Beyond this dialectal contrast, the articulatory features of both realizations of intervocalic /t/, especially in more extreme singing conditions, may also influence the final articulatory outcome. Although based on distinct analyses, this thesis adopts a multidisciplinary approach which aims to integrate both methodologies into a detailed and holistic account of sung word-internal intervocalic /t/ pronunciation. This specific pronunciation trait is approached from diverse theoretical and experimental perspectives, from social-indexical to articulatory considerations.

This thesis is divided into three parts. In Part I, I review the existing literature to provide the theoretical background for this work and situate it in a broader, multi-

disciplinary research context. Accounting for British singers' /t/ pronunciation must draw on research in fields including sociolinguistics, articulatory, acoustic, and experimental phonetics, sociology, popular music studies, and musicology. Each chapter in this Part I gives an overview of a relevant theoretical approach, which is inevitably not exhaustive. Chapter 1 starts with a general introduction to sociolinguistic variation, looking specifically at the concepts of style and style-shift. The main theories of style-shift in variationist sociolinguistics are delineated. The chapter ends by presenting pop culture and popular music more specifically as legitimate objects of study for sociolinguistics. In Chapter 2, I focus on the field of the sociophonetics of popular music, starting with an overview of existing works, from Trudgill (1983)'s pioneering study to recent developments. I tie in the theoretical aspects of style-shift seen in the previous chapter with concrete examples from popular music, and contextualize the phenomenon historically. In Chapter 3, I examine the singing voice in more detail, looking at its specificities and characteristics. An articulatory phonetic framework on which this thesis is based is reviewed, with special attention to the principle of economy of speech gestures and vocal effort minimization. The chapter ends by linking this articulatory account to the singing voice, through the notion of *singability*. Chapter 4 considers the phonetic feature studied in this thesis, the pronunciation of word-internal intervocalic /t/. Acoustic, articulatory, and sociophonetic accounts of intervocalic /t/ are explored and used to develop a hypothesis that singing, especially in extreme conditions, will result in British singers increasingly favoring a flap /t/ realization. Part I ends with an overview of the motivations for the present thesis, as well as the main research questions to arise from this wide-ranging literature review.

In Part II, I examine the modification of British singers' linguistic behavior using classic sociocultural perspectives, through an extensive study of real-world examples taken from traditional British heavy metal and British pop music from the 1980s. Chapter 5 focuses on the pronunciation of four archetypal traditional British heavy metal bands throughout their careers, connecting the analysis to relevant theoretical frameworks and incorporating the artists' own stated attitudes and perspectives to shed further light on the subject. Chapter 6 carries out a similar analysis, focusing on a broader sample of pop music recordings. Though the majority of this thesis is focused on the realization of /t/, Chapter 7 provides a smaller-scale comparative acoustic study of several vowel features for a subset of artists from the heavy metal and pop corpora. The last chapter in Part II, Chapter 8 explores the perception of intervocalic /t/ variants and how it has evolved for speakers of British English over time.

Part III applies experimental methods to test the articulatory hypothesis of Part I,

using lip tracking, electropalatography (EPG), and MRI imaging techniques. Five British popular music singers contributed to these production studies. The first experiment, involving lip tracking to measure lip aperture differences across singing and speaking conditions, is described in Chapter 9. Chapter 10 is dedicated to the main technique of this study, namely electropalatography (EPG). It looks at linguopalatal contact and how it may be influenced by factors such as pitch or syllable duration, which in turn may impact /t/ realization. The last experimental approach is considered in Chapter 11, with a more descriptive perspective of overall vocal tract configuration using real-time MRI. I end this thesis with a General Conclusion, in which a general discussion of the results is presented, as well as their possible theoretical implications and future research directions.

Part I

Background

1 Sociolinguistic Variation

Before any attempt to examine the tendency for singers to modify their pronunciation when singing, it is necessary to first look at the broader question of how pronunciation fluctuates in speech. It is clear that no two individuals speak in the same manner, but more than that, there is significant variation within individuals across different times and contexts. This chapter introduces a variationist sociolinguistic approach to language as the theoretical foundation underlying this work. Key concepts in the evolution of this discipline are first presented, with particular emphasis on the notion of style, and on the way stylistic variation has been applied to the realm of popular culture specifically. Variationist sociolinguistics has been defined as "the relationship between language and society" (Trudgill, 2000, p. 21), "the study of systematic and inherent variation in language" (Tagliamonte, 2011, p. xvi) or the observation of "language use in natural social settings and categorizing the linguistic variants according to their social distribution" (Chambers, 2013, p. 1). These approaches share the aim to identify, study, and account for aspects of language that vary according to aspects of society and social dynamics. It is primarily empirical in its method and quantitative in its approach to the study of language variation. A specific feature is chosen by researchers and its variants are identified and quantified in a single speaker or in groups of speakers. These features can be found at different levels of language, from vocabulary (*I gave him a right dirty look*; *I gave him a really dirty look*) to syntax (*She were a good worker; She was a good worker*) or pronunciation (*We were having a good time in what we were doin'*)[1]. Though variation occurs at all levels of language, this section and this work more generally will refer mostly to examples from speech.

[1] These examples are taken from Tagliamonte (2006, p. 10-11), themselves found in the York English Corpus.

A foundational study in sociolinguistic variation is Labov's doctoral thesis, "The Social Stratification of English in New York City" (1966), specifically his pilot study on the realization of "post-vocalic (r)"[2] in New York department stores (Labov, 1972). He visited three different department stores, ranging from a budget store frequented by working-class customers to an expensive luxury store frequented by the wealthy, and asked which floor he was on. This would elicit a casual "fourth floor" response from the employees, followed by a more careful and emphatic response when he asked employees to repeat themselves. This allowed him to identify and quantify realizations of post-vocalic /r/ in different social contexts, and to reach the conclusion that this specific feature is an indicator of social stratification in New York. Overall, the employees with higher socioeconomic status were more frequently rhotic in their pronunciation than the employees with lower socioeconomic status. In this research, Labov introduced one of the key constructs of variationist linguistics, the sociolinguistic variable. The sociolinguistic variable is an aspect of language "correlated with some non-linguistic variable of the social context: of the speaker, the addressee, the audience, the setting, etc." (Labov, 1972, p. 237). In other words, the different ways in which speakers pronounce a certain feature (in Labov's case, post-vocalic /r/) can be correlated with different broad social categories (here, real or perceived socioeconomic status). However, since Labov, there have been major developments in the field of variationist sociolinguistics, taking into account the relation between linguistic variables and more complex social and identity characteristics.

1.1 The Three Waves of sociolinguistic variation

This section examines the evolution of concepts and paradigms throughout the later development of variationist approaches, drawing considerably from Eckert's work on the Three Waves of Variation (2012, 2018). Central to this overview is the concept of indexicality as defined by Silverstein (2003) and used throughout this thesis as a way to describe the link between language variation and social meaning. Indexicality is concerned with how linguistic elements, through their use, point to or reflect social context, identity, and interpersonal relationships. In other words, these elements provide a way of capturing the social through language, with a specific feature of language associated with, and representative of, a certain population or category. The Three Waves of Variation perspective presents the way indexicality has been used in succeeding analytical positions.

[2]Labov uses the notation (r), but I will henceforth refer to it as /r/ as it is a more modern convention.

The First Wave: predetermined broad social categories

The first wave of sociolinguistic studies began with Labov's (1966) foundational work, as described in the introduction to this chapter. Variations of linguistic forms are linked with preimposed, often very broad social categories. Gender, social class, age or geographical origins are the macro-categories used to identify and describe patterns of linguistic variation, and speakers are classified accordingly. In other words, in the first wave, "speakers emerged as human tokens–bundles of demographic characteristics" (Eckert, 2012, p. 88). Linguistic variables are used to index speakers' membership to certain categories. While it is at the basis of variationist sociolinguistics and still used in current works, this approach is often criticized for being too 'top-down' and not concerned enough with the specificities of speakers themselves.

The Second Wave: local social dynamics and networks

Responding to perceived shortcomings in foundational sociolinguistic research, the second wave uses methods from ethnographic research to study inter-speaker variation through the local lens of relationships and social networks. Second wave researchers are interested in how participation in different kinds of communities (such as living in the same neighborhood or going to the same school) impacts sociolinguistic variation. The tighter-knit the network, the more norm-enforcing it is: engaging with and belonging to a group actively serves to maintain the linguistic norms present in said group. This idea was developed in Milroy (1980), who looked at phonological variation in three working-class Belfast communities. Similarly, Eckert (1989) conducted an ethnographic study of teenagers in high schools in the Detroit area. She identified two groups, the 'jocks' and the 'burnouts', where 'jocks' as described as a school-oriented, middle class community of practice, and 'burnouts' represent a locally-oriented, working-class community. Their distinct behaviors, social norms, and attitudes toward school were observed to account for their linguistic use. She found, for example, that burnouts were more strongly associated with the use of non-standard forms of speech such as negation. Unlike the first wave's macro-categories, the second wave highlights the fact that linguistic variables can index more narrow characteristics, and serve as resources in building interpersonal identities and connections.

The Third Wave: speakers as stylistic agents

While the second wave refines the macrosociological approach of the first wave and opens group-belonging as a potential mechanism, both perspectives focus on speakers as members of a specific category, reflected in their language use. The third wave is preoccupied with the speaker as an agent rather than merely a passive representative of a certain category, identified with a social group. As such, the third wave operates a reversal in the way the link between language and society is considered, where speakers are no longer passive carriers of dialect features, but use such features in different ways and at different times in order to reflect and construct certain identities or position themselves within or against a certain group. In the third wave, as Eckert puts it, "patterns of variation do not simply unfold from the speaker's structural position in a system of production, but are part of the active–stylistic–production of social differentiation" (2012, p. 98). Building on indexicality, Eckert invokes the concept of enregisterment, whereby "distinct forms of speech come to be socially recognized (or enregistered) as indexical of speaker attributes by a population" (Agha, 2005, p. 38). Enregisterment is the outcome of indexicality as a process in which linguistic features acquire contextual social meaning, and are used to indicate context-dependent characteristics or identities. In more concrete terms, a study by Zhang (2005) showed how certain Mandarin features have become enregistered as indexical of a new class of Chinese 'yuppies' in Beijing. Most notable is their use of the full tone, associated with Hong Kong and Taiwan and absent in Beijing. Through this, the yuppies mark their belonging to a more global, international sphere and differentiate themselves from the more traditional and local state managers. Another example is the work on variation of meaning in the voice quality of Heath, a medical student, in work and social contexts (Podesva, 2007). His use of falsetto is more frequent and longer in duration when at a barbecue with friends, where he adopts a more flamboyant, socially expressive gay identity. This is something he does not perform in a clinical setting where his speech expresses a more professional identity. These studies, and the third wave approach more generally, are concerned with the notion of speaker agency and the concept of style, which serves as a way to create and invoke different identities in response to specific contexts. Section 1.2 below discusses style in more detail.

1.2 Style

When attempting to define style, Coupland (2007, p. 2) first mentions literary style, an aesthetic concept with which most are familiar. Much as literary style refers to an author's distinctive way of writing, style in sociolinguistics refers to a distinctive

way of speaking. As variationist linguistics reconsidered the role of social groups on language variation, it placed increased focus on the notion of style as intra-speaker variation, whereby individuals adopt different linguistic features according to context as a tool for identity-construction. Accordingly, stylistic choices are inseparable from social meaning: choosing, actively or not, to say something a certain way as opposed to another can be seen as a way of creating social belonging and/or differentiation. In other words, style is what a speaker does with language in relation to others, and derives meaning from the association of linguistic features with particular social groups. Furthermore, style is something that occurs on a moment-to-moment basis, and is subject to change at any given time. Bell (2002) makes these dimensions of style salient in a simple question, which he sees as the core interrogation of any research on the topic: "*Why* did *this speaker* say it *this way* on *this occasion*?" (Bell, 2002, p. 139, original emphasis). The dynamic dimension of style, evident in this quote, is discussed more thoroughly in the following section.

1.2.1 Style-shift

Inherent to the concept of style in sociolinguistics is stylistic variation, or style-shift, as "there are no single-style speakers" (Labov, 1972, p. 208). Style is not static but continuously changing and evolving with the context of interaction in which a speaker is placed. It should therefore be recognized that in a specific context, speakers only show a portion of their full linguistic range of style. Certain variables might remain consistent across different styles, but every speaker has a set of language features that vary depending on the situation.

Style and style-shifting were first defined by Labov in his Attention to Speech Model (1972). Based on his observations of sociolinguistic interviews, he suggested that "styles can be ranged along a single dimension, measured by the amount of attention paid to speech" (Labov, 1972, p. 208). Labov saw style as a continuum from careful to casual speech, shifting according to certain triggers. As such, when talking about childhood habits or dangerous situations, or when communicating with loved ones, people tend to use more casual speech as they pay less attention to how they are speaking. On the other hand, in certain guarded situations, an individual may produce more self-conscious speech and therefore shift to a more formal style, as would be the case when reading aloud passages from a text or lists of words.

Though it did put the idea of intra-speaker variation in the spotlight, this approach has been criticized because of its limited, unidimensional view of style in terms of attention paid to speech alone. More styles exist than can be found on continuums of careful or formal versus casual or informal speech. Some have objected that reading

and speaking styles cannot be analyzed on comparable grounds (Milroy, 1987), or that it is in fact possible for speakers to be aware of their speech and still shift to a more casual style, sometimes even highly exaggerated (Schilling-Estes, 1998). Taking a more general approach to style than Labov's, Schilling-Estes sees stylistic variation as "shifts in and out of different language varieties, and shifts in usage levels for features associated with these varieties." (Schilling-Estes, 2004, p. 376). Characteristics of style-shift are put forward not only in terms of attention (conscious versus subconscious shifts); but also of duration (long-standing versus very brief shifts); indexicality (where specific patterns of stylistic variation are linked to a specific group), and organizational level (phonological, syntactic, semantic shifts etc.). Different types of variation can be identified and described:

- **Dialect shifts**, which involve variation in speech habit characteristic of a precise geographical area, time period, or social group;
- **Register shifts**, which link back to Labov's (1972) distinction between casual versus formal speech;
- **Genre shifts**, which involve shifts into and out of highly ritualized, routinized varieties such as 'legalese' or Legal English (Mellinkoff, 2004) used by lawyers, for example;
- **Code-switching**, as first conceptualized by Gumperz (1982), which is a specific type of shift involving switching between different languages as opposed to different varieties of the same language.

Considering the wide range of forms style-shifting can take, experts in the field have attempted to theorize the notion of stylistic variation through various approaches. The main ones are discussed in the Section 1.2.2 below.

1.2.2 Sociolinguistic theories of style

Giles (1973): Speech Accommodation Theory

The premise of Speech Accommodation Theory (SAT) first proposed by Giles (1973) is that an individual's speech patterns are dependent on who they are talking to. With origins in social psychology, and based on research on similarity and attraction, SAT puts the relational aspect of interaction at the center of its approach. Variation in speech is therefore studied in light of the addressee, where speakers are expected to adjust, or accommodate, their style toward or away from the person they are addressing. Generally, a speaker will tend to converge toward their addressee's speech style as a way to gain approval and facilitate communication. There may also be occasions where

the speaker will deliberately diverge from the addressee's speech style, as a way to accentuate differences and underline the distance between interlocutors. For example, one study found that American TV host Larry King would tend to match his style to his most prestigious guests, such as then-President Bill Clinton, whereas lower-status guests converged to King's speech style (Gregory & Webster, 1996). Later, SAT was reworked and renamed Communication Accommodation Theory (Giles et al., 1987) to match the extension of its scope to include broader social and relational dynamics of interaction.

Bell (1984, 2002): Audience Design

Bell's Audience Design has become a highly influential perspective in the study of style. The idea originated from the study of a range of sociolinguistic variables in radio news broadcasts in New Zealand. It so happened that two different radio stations in the same studio had the same newsreader working for them, allowing Bell to compare his speech styles in two different broadcasting contexts. He found that the journalist's style used more standard forms (such as the pronunciation of intervocalic /t/ as [t]) when he was on national radio, and conversely more colloquial forms (in the case of intervocalic /t/, the tapped allophone [ɾ]), when reading news on the community station. All factors are constant across the two situations (same speaker, same context, same source, etc.), except the audiences of the two stations. Therefore it must be that the audience factor is the driving force behind the speaker's shift in speech style. From this observation, Bell builds his framework of Audience Design, which can be summarized by the idea that style-shift occurs primarily in response to the speaker's addressees. In other words, speakers primarily design their styles for and in response to their audience. At its core, this framework closely resembles SAT, but places itself in a quantitative Labovian perspective, as opposed to Giles' (1973) social psychology approach. Figure 1.1 illustrates the relationship between the concepts Bell uses to characterize style design, which are discussed below.

Bell distinguishes between two types of style-shift, the first being *responsive* style-shift. As suggested by the name, responsive style-shift occurs in response to a specific situation, more specifically the audience of that situation. Thus, within the scope of what is referred to as *audience design*, the main source of influence to which a speaker responds is the second person addressee, the most direct audience member. Bell also mentions that speakers can attune to other, less important members of an audience such as *auditors* or *overhearers*, who are classified based on whether they are known, ratified, or addressed by the speaker. These third persons, however, affect style to a lesser degree than the second person addressee. Additional non-audience factors such

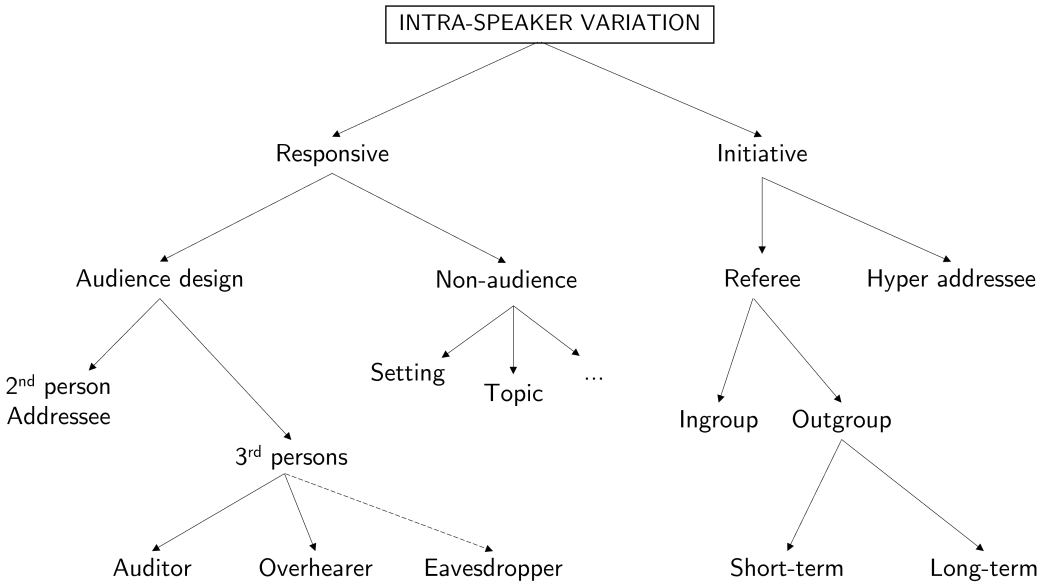


Figure 1.1: *Style as audience and referee design according to Bell (1984) and reworked in Bell (2002).*

as topic or setting may also influence a speaker's linguistic choices, though with less effect than the audience itself.

The second type of style-shift delineated in Bell's model is what he calls *initiative* style-shift. This initiative dimension of style is less developed in Bell's original Audience Design framework (Bell, 1984), which has led him to his reworking of the concept at a later date (Bell, 2002). While audience design is responsive, and therefore occurs in answer to a change in situation, the initiative type of shift is where style is used to redefine a situation, thereby initiating the change. This opposition is also discussed in terms of reactive and proactive style-shift in Schilling-Estes (1998), looking at performance speech. The primary type of initiative style is *referee* design. Referees are defined as "third persons not physically present at an interaction but possessing such salience for a speaker that they influence speech even in their absence," (Bell, 1984, p. 186). In other words, even though referees may not be physically in attendance during the interaction, they serve as models of reference for speakers. Speakers shift their communication style to align with the referees, rather than their immediate addressee, as a way of expressing identification with them. Referee design can be understood further through the distinction between the *outgroup* and *ingroup* dimensions. Ingroup referee design happens when speakers are not part of the group they are addressing, and shift toward their absent ingroup in a deliberate attempt to mark the difference between themselves and their direct addressees. For instance, a bilingual might use French to address a monolingual English speaker in Montreal, a city where both French and English are typically spoken (Giles, 1980). On the other hand, outgroup referee

design refers to a situation when speakers will diverge from the speech of their ingroup toward an outgroup which holds prestige or authority for them on some dimension. This type of outgroup referee design can be limited in time to a specific interaction, i.e. *short-term*, or work *long-term* and even become standard, as is the case for a speaker who has moved to another dialect region and accommodates over time to this non-native dialect. Lastly, and still within initiative style-shift, Bell also introduces the *hyper addressee* category, as a convergent type of style-shift where speakers shift toward the style of their direct addressee in a face-to-face interaction. For instance, Bell mentions Coupland (1980) studying the case of a travel assistant converging toward the speech of her client to insist on her helpfulness.

Bell's Audience Design is perhaps the most developed framework for understanding style-shift to date, but has faced criticism, in its original and reworked forms, after which initiative style-shift took on a more central role. The model's primary assumption is that style is essentially responsive, as "people do, after all, generally spend more time responding to others than taking the initiative" (Bell, 1984, p. 184). As such, the relative importance of the speaker's agency in stylistic variation as compared to speaker responsiveness is de-emphasized. Furthermore, the concept of audience can be called into question, as it may apply seamlessly to Bell's study of radio newsreaders, but is less clear in a direct face-to-face interaction between two interlocutors. In the continuity of the first iteration of Audience Design, the initiative dimension of style-shift and specifically its referee design aspect have been explored in more depth in another framework, that of Acts of Identity as defined by Le Page and Tabouret-Keller (1985).

Le Page and Tabouret-Keller (1985): The Acts of Identity Framework

The Acts of Identity framework is centered on the same process described by Bell's initiative referee design (Bell, 1984), though the focus moved from the importance of the audience to the importance of speakers' agency in identity-construction as the driving force behind style-shift. In a sense, this theory of style comes the closest to what Eckert associates with the Three Waves of Variation studies (Eckert, 2012) (see Section 1.1), where style becomes an active resource for the creation and re-creation of one's identity. The phrase *acts of identity* is first found in Labov (1963) which he uses to describe the fact that native inhabitants of the resort of Martha's Vineyard used specific linguistic features to distinguish themselves from summer visitors. Drawing from this observation and fieldwork conducted in the 1960s in the British Caribbean (Le Page, 1968), Le Page and Tabouret-Keller's Acts of Identity is summarized as follows:

> [T]he individual creates for himself the patterns of his linguistic behaviour so as to resemble those of the group or groups with which from time to time he wishes to be identified, or so as to be unlike those from whom he wishes to be distinguished. (Le Page & Tabouret-Keller, 1985, p. 181)

The idea that language and identity are inherently linked is central to this framework: language use allows speakers to project images of themselves, specific identities, at given moments. Speech acts, therefore, are considered acts of identity, as speakers externalize different representations of their identity at different times using stylistic variation. More precisely, speech acts are acts of identity *projection*–aimed outwardly at others. The feedback received from others reacting to this identity projection may lead to reinforcement. The speaker's style in that context may therefore become more *focussed*. In other words, to the extent that the speaker's projected identity coincides with the projected identities of the group, their linguistic behaviors are strengthened and become more sharply defined. If, conversely, the image projected through linguistic behavior does not match that of the group, it is said to be *diffused*. In any event, modifying one's linguistic behavior to match a specific group with which one wishes to identify is ruled by four riders, as described in Le Page (1968), Le Page et al. (1974) and Le Page and Tabouret-Keller (1985). As such, stylistic variation as an act of identity is constrained by:

1. **The extent to which the speaker is able to identify the model group**. Some groups, such as the family, may be more easily distinguishable than others, the latter even just being a vague notion of a group seen as the 'other'.

2. **The extent to which the speaker has access to the model group** and is able to formulate the rules of their linguistic behavior. Speakers need to know the specific linguistic behaviors of the group they are trying to identify with, which may be difficult if they have little access to that model group. Mistakes over the model group's speech or forms of hypercorrection may occur.

3. **The strength of the speaker's motivation** in imitating a model group, which may come in conflict with other model groups or the speaker's desire to retain their unique identity. Since identities fluctuate according to context, they can come in conflict with one another and a speaker's linguistic behavior may change accordingly. For instance, an individual may wish to identify with an economically successful group while maintaining a sense of his or her own native identity (Le Page, 1986).

4. **The ability for the speaker to actually modify their linguistic behavior**. This may lessen with age for example, though more experience can also lead to better proficiency.

To summarize the Acts of Identity framework, Le Page's image of speakers as "linguistic chameleons depending on the identity [they] are trying to project in any particular context" (Le Page, 2017, p. 18) is a useful summary. Speaking more broadly, this approach establishes two key points which can be equated to the main concerns of style-shift research associated with the third wave perspective described in Section 1.1. The first is that speakers do not simply adjust their style in response to the immediate speech environment (such as audience or level of formality), but instead actively and creatively engage with stylistic resources. Additionally, speakers are not necessarily constrained by external factors when shaping their speech and instead use language to influence and redefine both the external context as well as their interpersonal and personal identities.

Despite differences and limitations, the three theories described in the present section are crucial to the thorough study of style and style-shift. Their presentation in this chapter has necessarily been general, and they will be re-examined in Chapter 2 in relation to the specific case of popular music singing. Before this investigation into the sociophonetics of the singing voice, the value of pop culture as an object of sociolinguistic and variationist study must be examined.

1.3 Sociolinguistic variation and pop culture

1.3.1 Pop culture: an object of sociolinguistic study

Defining pop culture

Providing a definition of pop culture (PC) is not as straightforward as it appears, though everyone has an idea of what it is. Perhaps the best way to introduce the concept is to consider its history. Pop culture emerged in the 1920s in the United States (Danesi, 2007), primarily as a way for young people to challenge and break away from strict social norms and traditions. This societal shift was expressed mainly through new music, fashion and overall lifestyles. Pulp fiction, magazines, popular movies, jazz music, radio broadcasts and the iconic image of a new modern woman, the 'flapper', all contributed to the rise of the phenomenon: a type of culture developed 'by and for the people' (Danesi, 2007). More broadly, this was facilitated by new technologies (from the gramophone to the radio and TV) and the business interests and opportunities

they generated. This commercial opportunity in turn established and strengthened the connection between pop culture, mass media, and the marketplace (Danesi, 2020).

Culture in itself is a broad and heterogeneous concept whose definition is outside the scope of this study. Even so, a distinction must be drawn between *pop culture* and *popular culture*. Often used interchangeably, popular culture is a broader term which refers to cultural phenomena across time periods and societies, as collective practices, beliefs and objects, shared by a specific population. Pop culture can be seen as a subset of popular culture, associated with the modern era specifically. Attempts at defining PC usually insist on several points, starting with what it is not: pop culture is often seen as 'low' culture as opposed to a more elitist, less accessible 'high' culture (Storey, 2021). It is also conceptualized in opposition to 'folk' culture, which implicitly involves notions of historical tradition and authenticity, not necessarily central to pop culture. These aspects can be found in the following definition by Weinstock (2021), where pop culture is described as

> social practices and activities in which people can engage without significant training, education or cost. Typically associated with youth culture and quick to morph or fade, pop culture practices and activities are often construed as forms of leisure that allow for personal expression, afford pleasure, and create community. Such practices and activities often ignore or display an irreverence toward tradition, established standards for 'quality' and 'good taste,' age appropriateness and those who seek to enforce these things. (Weinstock, 2021, p. 20)

Another thing to consider is the sheer extent of pop culture. Since its beginning, PC has been linked to mass media, mass distribution and mass consumption, making widespread commercialization, mediatization and overall economic considerations central characteristics (Danesi, 2007; Werner, 2018). In this work, as has been done so far, the term pop culture will be preferred over popular culture as the focus is put on the modern, mass-media-driven cultural phenomenon and how it intersects with sociolinguistics.

Pop culture and sociolinguistics

The wide scope of pop culture, reaching most if not all aspects of daily social, cultural and even political life, makes it of interest to a number of academic fields, ranging from cultural, media and communication studies to literary or philosophical accounts. Since the 1980s, PC has established itself as an object of study in sociolinguistic research. Interest in pop culture is widespread among linguists, on account of "the obvious fact

that it is a multimodal phenomenon with a strong linguistic component" (Werner, 2022). Many expressions of pop culture involve some kind of linguistic feature, whether it be written or spoken.

There are two main areas of research in the sociolinguistics of pop culture: the influence of media on language change, and the sociolinguistics of performed language or performance–the latter being central to this work. The first approach to pop culture in sociolinguistics uses concepts of media communication and mediatization as tools for analyzing the interrelations between sociocultural and media-communicative change. The variationist sociolinguistic approach thus conceptualizes pop culture and mass media, very often through the example of television, as "a contextual parameter or independent variable whose impact on language variation and change can be studied in terms of single sociolinguistic variables" (Androutsopoulos, 2014, p. 13). Though it has been argued that, since TV does not provide face-to-face interaction, it can have little impact on linguistic innovation and change (Trudgill, 1986, 2014), evidence has also been found that engaging with TV can be a factor of language change. In a study on TH-Fronting in teenagers from Glasgow, Stuart-Smith et al. (2013) found that besides factors such as dialect contact, strong engagement with the London-based TV show *EastEnders* was also an indicator of language change in this community.

The other widespread approach to pop culture research in sociolinguistics is through the lens of performance. Bell and Gibson (2011) define sociolinguistic performance as a form of language use that deviates from the ordinary, crafted to appear as a deliberately put on or *performed*. A distinction is thus made between everyday performance and *staged performance*, "the overt, scheduled identification and elevation (usually literally) of one or more people to perform, typically on a stage or in a stage-like area such as the space in front of a camera or microphone" (Bell & Gibson, 2011, p. 557). Though mediated, TV shows, movies, radio shows, and popular music recordings are all instances of staged performance. If performance holds an important place in sociolinguistics research, it is because exaggerated linguistic features and usage is central to an act, with specific forms associated with specific genres. A more detailed examination of the sociolinguistics of performance within popular music is provided in Section 1.3.2 below.

1.3.2 Sociolinguistics and popular music

Defining popular music

Popular music is central to pop culture, which "may never have materialized in the first place without music to propel it into the social spotlight." (Danesi, 2007, p. 206).

Though the concept of popular music, much like pop culture, is difficult to define, four types of meanings popular music can take have been identified by Birrer (1985) and are still used by many to refer to the concept (see for example Middleton (1990) or Frith (2001)):

- **Normative definition**: popular music is an inferior type of music;
- **Negative definition**: popular music is music that is not something else. It is usually opposed to 'folk' or 'art' music;
- **Sociological definition**: popular music is associated with (produced for or by) a particular social group;
- **Technologico-economic definition**: popular music is disseminated by mass media and/or a mass market.

These aspects of popular music are effectively the same as those used to define pop culture, confirming Danesi's view that "pop music and pop culture are virtual synonyms" (Danesi, 2007, p. 206).

Much like for pop culture, a caveat is in order on the differences of meaning in qualifying music as *pop* or *popular*. In cultural studies, there does not seem to be unanimous agreement on the exact terminology. The two terms are sometimes used interchangeably (Danesi, 2007), or no justification is offered for choosing one term over the other (for example, Adorno (1990)), while at other times academics insist on a clear distinction between *pop music* and *popular music* (Frith, 2001). In this work, I will refer to popular music as an umbrella term which "includes all contemporary popular forms–rock, country, reggae, rap, and so on" (Frith, 2001, p. 94). In other words, popular music can be taken as a broad category which encompasses all non 'art' and 'folk' genres. On the other hand, pop music will be used when referring to the specific genre of pop, as "what's left when all other forms of popular music are stripped away" (Frith, 2001, p. 95), insisting on the restrictive aspect of pop as its own musical genre, different from rock, country, reggae etc. Music genres can sometimes be difficult to categorize, but at its core the idea is "that a genre requires a certain sound, which is produced according to conventions of compositions, instrumentation, and performance." (Weinstein, 2000, p. 6). Heavy metal and pop being the two main genres covered in this thesis, their respective codes are described in their dedicated chapters, Chapter 5 for heavy metal and Chapter 6 for pop.

The sociolinguistics of performance

In the restricted field of the sociolinguistics of performance, popular music is generally considered a prime example of how linguistic stylization in pop culture may be used as a

tool for expressing identity and social meaning. In their defining paper on the subject, Bell and Gibson (2011) draw attention to five dimensions of language performance, described below.

1. **Identity**: As a starting point, identity is defined as both the product of where individuals come from and the sum of their unique lived experience, as well as the "process of something negotiated and constructed rather than just being there" (Bell & Gibson, 2011, p. 561). Bell and Gibson underline the importance of this concept in the sociolinguistics of performance, where artists 'doing' a particular accent are *referring* (see Section 1.2.2) to the accent and therefore referencing a number of the characteristics of a sociocultural group. This is usually one that is different from the singer's own group, and that the audience can recognize.

2. **Reflexivity**: Performance is taken as a reflexive activity, which directs attention on itself. In terms of linguistics, reflexivity in performance can be seen through hyperbolic forms and the use of phonetic features such as vowel lengthening, or exaggeration of prominent features.

3. **The audience**: The audience (real or targeted) is another dimension that is crucial to performance, and which distinguishes it from other linguistic activities. Performance is not only directed at an audience, but is also created *for* them (Coupland, 2007). Audiences, whether they are live or distanced in the case of recorded or mediated performances, influence the artist. For live performances, this is done on an immediate basis through the direct feedback of the audience. For mediated performances, the artist always has to keep in mind the existence of an audience, whether real or imaginary, and anticipate its potential response. Additionally, a more long-term relationship between audience and performer is also created through the expectation of repetition: though artists may create new material over the years, they are usually expected to replay fan favorites in their performances.

4. **Authenticity**: Artists are not necessarily concerned with authenticity in the sociolinguistic sense of one's vernacular being their most genuine speech. This is because of the intentional stylization of language as a basis of performance. Nevertheless, depending on the performance or the genre, authenticity can also play a central role. In folk music, for example, the artists' persona and their genuine self is often taken to be one and the same (Coupland, 2011). Sometimes, a specific accent can also be expected for a specific type of music, in order to mark its authenticity.

5. **Genre**: Bell and Gibson (2011) refer to genre in its broad acceptation, distinguishing, for example, film from theater performance from stand-up comedy. Followed in this work is Coupland (2011)'s more restrained, specifically musical approach to genre as a distinctive form of music with its own performance codes, such as rock, hip hop or country. As such, this is directly linked to the notion of authenticity–a certain type of linguistic behavior is expected for specific genres.

Bell and Gibson (2011) also insist on other modalities of performance that are not directly linked to language. The music itself, but also the set, appearance, and gestures linked to an artist's performance are also important dimensions to consider. Pop culture and its expression in popular music offers a rich field for sociolinguistic exploration. Its multimodal nature, wide-reaching influence, and role in shaping and performing identity, provide opportunities to understand how language shapes and is shaped by social and cultural dynamics, and therefore points to social meaning.

1.4 Summary and conclusion

This chapter was intended as a broad introduction to the field of sociolinguistics and its approach to variation. It aimed to outline important concepts and paradigms that will be used as the theoretical foundation of this work. Following the Three Waves of Variation study (Eckert, 2012), indexicality and enregisterment were defined respectively as the process of language elements acquiring contextual social meaning, and the outcome of this process, where specific linguistic features become socially meaningful and associated with particular social registers or identities. The notions of speaker agency and style corresponding to the Third Wave
indexThree Waves of variation study were described in more detail, and the main sociolinguistic theories of style, Speech Accommodation Theory (SAT), Audience Design, and Acts of Identity, relevant to this work were reviewed. Finally, both pop culture and popular music were defined and shown to be relevant objects of sociolinguistic variation study. The last section ends on an overview of the sociolinguistics of performance as construed by Bell and Gibson (2011) and Coupland (2011), more specifically in terms of musical performances and singing.

This chapter attempted to show how language variation is deeply embedded in social meaning and identity construction, and that pop culture and popular music performance provide a unique perspective in the field of sociolinguistics. Though the approach in this chapter was general, it sets the theoretical foundations for a closer look at the sociophonetic aspects of the singing voice, which is the focus of Chapter 2.

2 The Sociophonetics of popular music singing

STANDING "at the intersection of sociolinguistics and phonetics" (Foulkes, 2006, p. 495), sociophonetics bridges the gap between these two fields through its focus on speech variation. Phonetics as a discipline has been at the heart of sociolinguistics since its origins in the work of Labov (1972), who identified intra-speaker variation that carries social-indexical information as an important area of research. Style-shift can be seen in a variety of contexts, and Section 1.3 introduced the study of language use in pop culture (PC) and more specifically popular music performance, as a relevant research target for investigations.

"'Accent' is perhaps the most obvious linguistic means by which performances index identity" (Bell & Gibson, 2011, p. 570) and accents provide a rich source of social meaning. It is no wonder, then, that a whole sub-field focused on the sociophonetics of popular music singing has developed from the linguistic approach to PC and performance. This chapter reviews the main research from this discipline, starting with an overview of Trudgill (1983)'s pioneering study on British pop song pronunciation, before outlining subsequent studies. The second part of this chapter addresses how existing theories of style-shift (delineated in Chapter 1, Section 1.2.2) can apply to the specific case of popular music singing. The typology of style-shift that occurs in popular music singing is discussed through the perspective of the evolution from 'Americanization' of the singing voice toward a recognized and standardized variety of English, specific to singing. Lastly, a chronology of main events in the evolution of popular music is provided in order to situate this work in its proper social and historical context.

2.1 Style-shift in popular music singing

2.1.1 An overview of Trudgill's *Sociolinguistics of British Pop-song Pronunciation*

Trudgill (1983)'s research on the sociolinguistics of British pop-song pronunciation is a pioneering study in the sociophonetics of popular music, and constitutes the foundation of this thesis and the broader literature. This section provides a general overview of this work, whose main ideas and concepts are developed in more detail throughout. Based on a mostly impressionistic analysis of pop music from the late 1950s and the early 1960s, Trudgill points to the fact that "singers of this form of music employ different accents when singing from when they are speaking" (p. 141). He identifies five distinctive features of this sung accent, which have since been adopted by most subsequent works on this topic. Subsequent research further conceptualized and categorized this system into the USA-5 model (Simpson, 1999). A description of the model can be found in Table 2.1. For clarity, features are named based on the corresponding phonological processes described by Wells (1982a).

Feature	Description
T-voicing	Intervocalic /t/ in words like *better* or *city* is pronounced as [d] rather than [t] or [ʔ].
Lack of TRAP-BATH split	The long open vowel [ɑː] in words like *dance* or *last* of Southern British English tends to be realized as a shorter and broader [a].
Rhoticity	Non-prevocalic /r/ is pronounced, even when the singer's regular speech is non-rhotic.
Monophthongization	The diphthong [aɪ] in words like *life* or *like* is realized as [aː].
LOT unrounding	Words like *body* and *top* are pronounced with [ɑ] rather than a rounded [ɒ].

Table 2.1: *The USA-5 model adapted from Simpson (1999)*

Although these were specifically identified as pronunciation tendencies in the singing of British pop music artists, these features also happen to correspond, as the name suggests, to the main differences between very broadly defined American and British accents. There are of course some British English varieties that do display some of these features, for example the lack of TRAP-BATH split being typical of the Northern varieties of British English, or rhoticity for Scottish English (Wells, 1982a). However, Trudgill points out that no single British variety presents all these features whereas they

are consistently found in American varieties, and moreover tend to be associated with a stereotypical American accent by speakers of British English. More importantly, most British singers who display these traits when singing do not do so when they speak. Any of these features can be described as Labovian sociolinguistic variables, where multiple ways of realizing the same pronunciation trait exist, varying according to the different aspects of a communicative situation.

Trudgill's explanation is that, since most pop music productions in the twentieth century were American, British artists leaned toward the use of American pronunciation features in order to identify with, or imitate, this model group:

> Americans have dominated the field, and cultural domination leads to imitation: it is appropriate to sound like an American when performing what is predominantly an American activity, and one attempts to model one's singing style on that of those who do it best and who one admires most. (Trudgill, 1983, p. 144)

Adding to his impressions of pop artists' pronunciation from the late 1950s to the early 1960s, Trudgill also took on a quantitative approach, comparing said pronunciation to that of a slightly later period, namely the mid to late 1970s. He noted a gradual weakening in the use of American-like features in bands that previously showed high percentages such as the Beatles, as well as mixed tendencies for pop bands such as Supertramp and punk bands like The Clash. He reached the conclusion that by the mid-1960s, the tendency to adopt features of the American model had significantly decreased, and posited that this change was due to the evolution of pop music itself, symbolized by the First British Invasion, which will be discussed in more detail in Section 2.3.2. British pop music consequently acquired validity with some British artists switching back to linguistic variants more consistent with their geographical origins. A clear example of that is the development of the punk genre, whose main audience were British working-class youths (Trudgill, 1983), and whose singers introduced new pronunciation features usually associated with 'low-status' Southern English accents. It is important to note that these artists did not entirely forgo an American-influenced pronunciation either, but rather both models competed against each other, in what Trudgill calls *acts of conflicting identity*, borrowing from the Acts of Identity theory's third rider (Le Page et al., 1974), described in Section 1.2.2. How Le Page's and other style-shifting theoretical frameworks apply to singing voice pronunciation is discussed in Section 2.2. Trudgill's study establishes the foundational link between British pop-song pronunciation and the sociocultural influence of the American model, later counterbalanced by a shift toward more local pronunciation.

2.1.2 After Trudgill

Many studies, building on Trudgill's foundations, have examined the phenomenon of style-shifting in popular music. A few relevant ones have considerably shaped and influenced the field, introducing new concepts and research paradigms. This subsection provides an overview of the most relevant and significant studies.

Simpson (1999): Mixed analytical methods and the importance of cultural context

Simpson (1999) was the first to follow up on Trudgill (1983)'s study, looking at the evolution of American-influenced realizations of British artists in later decades. He studied the pronunciation of several artists such as Dire Straits or Van Morrison, choosing excerpts from a variety of songs. He confirmed the tendency identified in Trudgill (1983) as a slight decrease in the influence of the American model, blending with the addition of new British English features. In addition to his direct findings, Simpson's work remains important as it extended and theorized the pronunciation features noted in Trudgill into the USA-5 model (see Table 2.1) still commonly referred to today in similar studies (see for example Konert-Panek, 2017a; Flanagan, 2019 or Campos-Astorkiza, 2024). He also introduced concepts from outside the field of sociolinguistics strictly speaking: from discourse analysis, insisting on the importance of field, mode, and tenor (Halliday, 1978), as well as from ethnomusicology and the sociology of pop music. In doing so, he anchored this type of research within the broader historical and cultural context of the evolution of the music industry. Quoting from popular music theorist Simon Frith, he highlighted the fact that "rock and pop musicians of whatever kind are acting according to *conventions*" (Frith, 1988, p. 4 quoted in Simpson, 1999, p. 344; original emphasis). In other words, artists' productions and thereby their pronunciation are always related to a set of norms or standards dictated by changing cultural patterns and influences.

Morrissey (2008): Audience, reference style, and the importance of phonological considerations

Drawing upon both Trudgill and Simpson, Morrissey (2008) discussed various factors which influence the tendency for style-shifting in popular music, specifically rock, pop and folk singing. Much like the previous studies, his data consisted of a small number of short (sometimes only a few words) lyrical excerpts from a small number of artists such as Lindisfarne or Ralph McTell. Morrissey looked at the ways different British

artists have adapted their pronunciation to a certain American model, while retaining some of their native British spoken features. He pointed out that "a crucial fact when we discuss style choice in popular music is that it is part of an act, a performance" (Morrissey, 2008, p. 195), and should be treated as such. This includes the idea that a distinction needs to be made between live performances and recordings, and that neither can be taken as an act of natural language. Because a performance is by definition aimed at someone (usually a group of people), it also means that an audience must be taken into account, which Morrissey does using the Audience Design framework (Bell, 1984). He argued that the different styles adopted in British pop music can be seen as *reference styles*. Certain genres of music such as blues, blues-derived rock and pop, or country are strongly rooted in American culture. It bears mentioning that it is more precisely African American culture and music which is at the heart of these genres. As put forward by Danesi (2007), "without African American artists and musicians, there would be no American culture as we know today" (Danesi, 2007, p. 275). This therefore creates expectations on the part of the audience, anticipating an American reference style. Conversely, traditional British folk songs have a reference style that is more oriented toward rural dialects of British English. Additionally, and perhaps the most relevant within the scope of Part III, Morrissey goes beyond sociolinguistic concepts and introduces phonological considerations inherent to the practice of singing: "as any vocalist appreciates, there are certain speech sounds that lend themselves better to singing than others" (p. 209). This idea is more thoroughly discussed in Section 3.2, where I address the specificity of the singing voice as an object of study.

Beal (2009): language-ideological approach and going beyond the American model

Beal's work investigated the use of specific dialect features in the singing productions of Northern English indie band Arctic Monkeys. While most previous works (including the ones discussed above) have focused on how bands tend to follow an American-inspired pronunciation model when singing, even when retaining some British features, Beal argued that in recent years, this American model has been abandoned and pronunciation style has shifted toward specific British regional dialects. In the case of Arctic Monkeys, Sheffield English replaced American English as the reference model of pronunciation–in other words, their musical productions are striking because they are singing in their 'own' accent. To explain this, Beal developed a new approach for this type of study, looking at style in popular music from a language-ideological perspective, using the notions of indexicality (Silverstein, 2003) and enregisterment (Agha, 2005), detailed in Section 1.1. The main idea put forward is that "in the context of

British popular music, regional pronunciations index authenticity and independence, whereas the 'American' features outlined by Trudgill (1983) index conformity to the pop mainstream" (p. 238). As an indie band, Arctic Monkeys presented themselves as independent of the global (American-dominated) music industry, rather insisting on local, northern working-class values. Consequently, using not only Sheffield-specific pronunciation features, such as [ʔ] for non-initial /t/ or [ʊ] for words belonging to Wells' STRUT lexical set (1982a), but also dialectal vocabulary and local references, Arctic Monkeys indexed values of 'local-ness' and authenticity. From a broader perspective, Beal's work explicitly articulated one of the main tensions underlying the studies discussed above–the central opposition between attempts to mimic and identify with an American model and the idea that an American-like pronunciation may be the norm in pop singing, without indexing an actual American identity. This tension between active 'Americanization' and a normalized pronunciation standard in singing is discussed in Section 2.3.1.

Coupland (2011): Broadening perspectives through semiotic elements of genre

Coupland (2011) extended the scope of the sociolinguistic study of popular music. He situated his research within diverse fields organized according to genre, in which the concepts of voice, place, and performance, must be reassessed. While acknowledging the interest in studying popular music from an 'Americanization' perspective, which has dominated since Trudgill (1983), Coupland suggested that this traditional sociolinguistic approach is too restrictive. Firstly, he argued that the understanding of place as 'territorial-indexical' (geography indexes a particular dialect) is too limited, and would be better apprehended through a wider cultural and acoustic epistemology of place. Similarly to Morrissey (2008), Coupland pointed out that equating voice with speech in popular music studies is problematic: "because speaking and singing are self-evidently distinct communicative modes, it may be misleading to bring speech variation perspectives to bear on singing" (Coupland, 2011, p. 575). He thus claimed that pronunciation or style in singing is just as likely to reflect specific phonatory constraints inherent to the practice, as it is social-indexical motivation. Another limitation of sociolinguistic studies identified by Coupland is that, although routinely acknowledged in some way, previous research does not take into account the distinctive fact that popular music needs to be approached and studied as a performance (whether live or mediated through a recording), a highly stylized situation that is organized according to a specific music genre and its codes. Revisiting concepts of place, voice, and performance, Coupland attempted to show how these elements may work together to constitute different genres

of popular music. The second part of his study applied these analytical tools to three songs by three artists representative of three broadly defined music genres: *Maybellene* by Chuck Berry for classic rock and roll; James Taylor's *Copperline* for folk/country and a cover of *Johnny B. Goode* by the Sex Pistols for punk rock.

Other studies

Various additional studies on style-shifting in popular music exist, most of which have drawn on the foundational works evoked above and attempted to approach the same questions from various perspectives. One way to classify these studies is to look at the speech characteristics of the artists and the targeted style model. In a majority of these studies, the usual focus is on British artists adopting or, in some cases, rejecting an American-inspired style. This is the case for Flanagan (2019), looking at how Arctic Monkeys have evolved since Beal (2009)'s study and their increasing fame, especially in the United States. A series of papers by Konert-Panek (2017a, 2017b, 2018) examine the 'Americanization' of several artists such as Amy Winehouse, Adele or Joe Elliott of Def Leppard, while Campos-Astorkiza (2024) focuses on the British Americana band Mumford and Sons. M. Krause and Smith (2017) analyze the style of two Glaswegian indie artists, and their respective attitudes toward an American target, broadening the perspectives of a purely phonetic analysis by including sociolinguistic interviews of the singers.

The influence of the United States is not only present for British artists, however. Australian country and hip hop singers also appear to position themselves within or against an American pronunciation model (O'Hanlon, 2006; Duncan, 2017), as well as artists from New Zealand (Gibson, 2019, 2023; Gibson & Bell, 2012). Though when it is targeted, American English is usually quite generic in its definition, African American English has also been studied as a more specific model for Australian hip-hop singer Iggy Azalea (Eberhardt & Freeman, 2015), and other non-US Blues singers (De Timmerman et al., 2024). Another interesting example is that of Rihanna, who was born in Barbados but whose career has been based in the U.S. Though she sings mostly in her 'own' American accent, she targets Caribbean English Creole in some songs, thereby reclaiming her Caribbean identity (Jansen & Westphal, 2017). A notable shortcoming in this research area, which this thesis does not avoid, is that studies are centered on popular music from the anglophone world, performed in English by people who are native English speakers. Nevertheless, a few studies have considered style-shifting in other languages, notably Westphal (2018) with German reggae singers targeting Jamaican Creole, Hayes (2023) identifying a pan-Latinx style in the Spanish-language performances of Puerto Rican and Colombian artists Bad Bunny and J Balvin,

and Owens (2023) with Quebec French singers of various genres.

The sociophonetics of popular music may be a small field in comparison with others, though detailed studies are available for a wide variety of music genres, artists, styles, and varieties of English. Despite different approaches, a common thread that runs throughout all of these studies is that of identity, and more specifically the tension between the overwhelming influence of the American model and the desire, in Beal (2009)'s words, "to go local." Before discussing this tension more closely in Section 2.3.1, I will first put these specific examples back into their theoretical context, looking at how various style-shifting paradigms may apply.

2.2 Theoretical implications

Most linguists working on style and style-shift have agreed that sung performance (and, in this case, by British singers) is a particularly good example of style-shift (Bell, 1984; Coupland, 2011), linked as it is to a specific communicative situation and obeying a certain number of strict sociocultural norms. This section therefore complements Section 1.2.2, relating various sociolinguistic theories of style-shift to the specific case of the sociophonetics of singing voice pronunciation.

2.2.1 Sociophonetic processes

In their paper on the sociolinguistics of performance, Bell and Gibson (2011) identified four kinds of sociophonetic processes that artists resort to in performances when shifting from one variety to another. All the studies previously discussed refer in one way or another to these processes, using them to analyze the singing pronunciation of the artists under consideration.

1. **Selectivity**: using some features of the target variety while omitting others.
2. **Mis-realization**: incorrectly realizing certain target features.
3. **Overshoot**: exaggerating the characteristics or frequency of features of the target variety.
4. **Undershoot**: minimizing the characteristics or frequency of features of the target variety.

These sociophonetic processes help explain how singers adjust their pronunciation to match different styles. In the specific case of British popular music singers, they can be used as a tool to analyze how these artists display the USA-5 features referenced in Table 2.1. Style-shift in singing, more complex than mere imitation, involves linguistic

factors as well as considerations of identity and audience expectations. The following section outlines how broader theories of style-shifting apply to singing specifically.

2.2.2 Theories of style applied to singing

Opening the theoretical discussion in Trudgill (1983), Speech Accommodation Theory (SAT) is an evident starting point for attempts to explain pronunciation modification by singers. SAT expects speakers to adjust their style toward (or away from) the person they are addressing. Though acknowledging its relevance, Trudgill (1983) rejects this model on the grounds that it is not sufficient to describe the phenomenon of pop-song pronunciation: it only applies to conversational situations which do not apply in the case of performance, and it seems difficult to argue that British pop singers are accommodating to their (primarily British) audience by adopting American-like features.

Better suited to explain pop-song pronunciation, according to Trudgill, is the theory of Acts of Identity as developed by Le Page (1968; 1974). This theory sees style-shift as an attempt by speakers to resemble a group with which they wish to identify from time to time. Based on the USA-5 model (Table 2.1) of pronunciation, this group is identified as "clearly, if somewhat loosely, ... 'Americans' "(Trudgill, 1983, p. 144). British popular music singers seem to be modifying their pronunciation toward an American model group with which they wish to identify when performing. From this perspective, the situational parameters of interaction type and audience that limited the scope of SAT are no longer a concern. Furthermore, Trudgill points out how all four of Le Page's riders described in Section 1.2.2 apply to pop-song pronunciation. Though not necessarily deliberate on their part, British pop singers seem to have successfully identified their model group as American, in keeping with Le Page's first rider. The second rider is concerned with the extent to which British singers have access to their model group and if they are actually able to work out their pronunciation rules. This may lead to cases of overshoot and hypercorrection, which can be equated to Bell and Gibson (2011)'s process of 'overshoot,' when speakers overapply a rule that is perceived as more 'correct' in environments where it is actually inappropriate. Trudgill mentions the use of such forms as proof that British singers have not sufficiently analyzed or assimilated the pronunciation rules of the targeted American model. This is the case with rhoticity–"singers know that, in order to 'sound like an American', one has to insert an /r/ [in non-prevocalic position] ... the problem is that some singers have not mastered the principle behind where this should and should not be done" (p. 148). For example, in the song "Bachelor Boy" Cliff Richard wrongly inserts an /r/ between the article *a* and the noun *bachelor*, leading him to repeat multiple times [ər bætʃələ]

in the lyrics "You'll be a bachelor boy." Hypercorrection or overshoot can also be related to the fourth rider, which concerns the ability to actually modify one's linguistic behavior. In addition, British singers' inconsistency in modifying their pronunciation should be highlighted. Despite being aware of the American features that should be adopted, British singers do not consistently pronounce them: for instance, despite British listeners agreeing that the Beatles 'sound American' on their first LP *Please Please Me* (1963), they only pronounce 47% of potential non-prevocalic /r/s (Trudgill, 1983). This can also be linked to processes of 'undershoot' or even 'mis-realization' mentioned in Bell and Gibson (2011). Finally, and accounting for its prominent position in both Trudgills' work and this section, the third rider of the Acts of Identity framework is probably the most relevant to style-shifting in popular music. It relates to the strength of various and possibly conflicting motivations toward one group or another, including the desire to retain their own unique identity. As highlighted by the title of Trudgill's study itself, 'acts of conflicting identity,' and seen in many of the other studies described in Section 2.1.2, a common thread is the existence of an unresolved overarching tension in singers between adopting the USA-5 model, taking on an American-like pronunciation, or trying to 'sound American' versus retaining a degree of their 'own' native accent, navigating the projection of different identities through their pronunciation. Overall, the Acts of Identity's four riders appear to adequately explain some of the underlying mechanisms behind singers' linguistic behaviors, namely how partial access leads to partial realization, while conflicting motivations are reflected in the projection of conflicting identities through accent and pronunciation.

In the first conceptualization of Audience Design, Bell discusses Trudgill's study, claiming that the adoption of American-like pronunciation features by British pop singers "bears all the linguistic hallmarks of outgroup referee design" (Bell, 1984, p. 194), as referenced in Figure 1.1 and delineated in the corresponding section. The American model acts as the linguistic referee that is external to the direct context (the United Kingdom). Though not perfect in their adoption of American pronunciation traits (Bell also mentions cases of mistakes and hypercorrection), the modification of certain features is enough to evoke for both singers and their audience the target pronunciation reference they are after. For Bell, the American model is part of the outgroup dimension of referee design, as British singers diverge from their own accent toward an outgroup which holds prestige and authority for them, at least in the field of popular music. Morrissey (2008) also views this phenomenon as a case of outgroup referee design, insisting on the fact that both the singers and their audiences implicitly agree that Americans are the referee, thus conceptualizing the collection of American-inspired pronunciation features as the *reference style* in popular music. Gibson and Bell

(2012) and later Gibson (2023) describe this interpretation more formally. Whereas, in Bell's model, outgroup referee design is associated with initiative style-shift–used to redefine a situation, with the speaker thereby initiating the change–the use of American features in singing is a special case which involves *responsive* referee design–where style-shift occurs in response to a specific situation. As such, in the case of popular music, the American-influenced referee design has become so deeply associated with the performance situation that it serves as the default style of reference. Even though it involves shifting away from the singers' own native spoken style, it is responsive "because of its predictability in and its appropriateness to the pop song context" (Gibson, 2023, p. 5). Trudgill (1983) touches on the notion of appropriateness as a relevant factor. He recognizes that since different situations, topics, and genres require different linguistic styles and registers, it is no more surprising that singers should sing in the appropriate style for popular music than vicars should preach in the appropriate style for church.

The three main theoretical approaches to style-shift, SAT, Acts of Identity and Audience Design, are useful to varying degrees when applied to the question of popular music singing. They share common ground in the fact that they all consider the phenomenon of linguistic modification a tool for identification. It seems, however, that from both these theoretical perspectives and their more concrete applications in the studies reviewed above, that a major question has yet to be definitively answered: whether this American-inspired shift is conscious and deliberate on the part of singers to imitate a targeted American model or if it is a standardized norm singers naturally default to.

2.3 Style-shifting in context

2.3.1 From 'Americanization' to Pop Song English

A chronological review of the main findings of studies on the sociophonetics of popular music presents a general movement away from Americanization as an act of identification with an American model, toward the interpretation that American-inspired pronunciation features are broadly recognized as a standard that singers may default to. Going back to Schilling-Estes (2004)'s typology of style-shift (see Section 1.2.1), this evolution raises a question: is the use of American traits in singing a case of dialect-shift–involving change in speech habit characteristics of a precise geographical area–or genre-shift–involving shifts in and out of highly ritualized and standardized varieties? From Trudgill's perspective, when British singers adopt American accents,

they do so as a way to align themselves with the cultural and musical dominance of the United States, the birthplace and center of popular music since it emerged in the 1920s (Danesi, 2007). While this may have been true at a given time–it bears reminding that Trudgill's work was conducted in the 1980s on music that came out mostly in the 1960s and 1970s–more recent studies have developed the idea that this American-inspired pronunciation style is simply the standard singing style in popular music, and that its use may less dialectically marked than previously thought. Morrissey (2008) and Coupland (2011) both touch on this idea when pointing to the fact that singers perform according to certain standards, whether in accordance to a certain reference style or the established norms of a specific music genre, respectively. These pre-established standards are ultimately what determine which variety of English is appropriate. A few studies have directly asked popular music singers about their pronunciation when singing and their personal attitude toward it (see Gibson and Bell (2012) and M. Krause and Smith (2017), for example). For instance, when interviewed, New Zealand-born singer Dylan Storey believes that "a NZ accent would 'sound ridiculous' in mainstream pop" (Gibson & Bell, 2012, p. 146). As discussed previously, Beal (2009)'s research on the Arctic Monkeys' use of their native Sheffield accent in their songs supports this idea and goes further: she argues that, because it has become the norm for pop music, using an American-influenced style is the default for artists, where singing in one's 'own' accent actually requires more conscious effort. Another singer interviewed in Gibson and Bell (2012) also corroborates this idea by admitting he had to re-record a song where New Zealand was central to the lyrics because his accent did not 'come through strongly' enough the first time. This suggests that while it is possible for singers to use their own accent in song, it requires more focus and effort. In that sense, the argument initially proposed by Trudgill (1983) is reversed: it is performing in one's native pronunciation that in reality requires a conscious and willful act of identity. Perception studies further support this, showing that British listeners expect an Americanized style in pop songs (Jansen, 2018), reinforcing its role as the norm for the genre.

If an American-influenced style is therefore the norm in popular music singing, the implication is that, though in the past there may have been a more conscious and overt American influence on the pronunciation of British singers, this American-influenced style has since become enregistered, or socially recognized as indexical (Agha, 2005) of pop singing, divesting it of some, if not most, of its geographical and dialectal connotations. Earlier American influence has evolved into what is considered an appropriate pronunciation style for pop singing, now a widespread standard adopted by British artists and others around the world. In Beal's words,

> While British pop singers in the 1960s and 1970s may have used Simpson's

> USA-5 model in their singing because these features were then indexed as American, in time the association of these features with a certain type of musical performance led to their being indexed in this context as "mainstream pop" rather than simply "American." (Beal, 2009, p. 229)

For this reason, it can be argued that this popular music pronunciation standard is in fact a variety of English in its own right, a phenomenon which Gibson (2023) refers to as Pop Song English (PSE). Pop Song English is defined as a global style of English used in popular music that is no longer linked to a specific country or origin. Instead, it serves as a standard way of singing that artists of diverse origins adopt to fit the expectations of global audiences. Over time, what started as a reflection of American influence has become a neutral feature of modern popular music, separate from its original association with the United States. To answer the question raised at the beginning of this section regarding the typology of this specific style-shift, it can be concluded that although both perspectives can be defended, the transition from dialect to genre shift is evident in the findings of sociophonetic research on popular music singing, and in the broader historical evolution of popular music.

2.3.2 A chronology of singing pronunciation style in popular music

The timeline of the evolution from a deliberate attempt to 'sound American' toward the enregisterment of this pronunciation as PSE is imprecise–it is difficult to imagine a single point in time when the switch from one to the other occurred, since it involved an interconnected group of artists, audiences, and intermediaries such as record labels. Most authors agree that 20th century popular music was first developed in and directly influenced by an 'American' model, while the shift toward a pop singing standard was finalized in the 21st century (Beal, 2009; Gibson, 2023; Trudgill, 1983; Westphal & Jansen, 2021). The 1980s, the period under examination in this work, thus may be understood as an intermediary moment in this evolution. On a cultural level, the decade was a major turning point in the production, distribution, and reception of popular music (Frith, 1988; Negus, 2011) and saw the creation of a 'new mainstream' in the U.S., with the return of the Top 40 radio format (based on a precisely curated and heavily repeated rotation of the most popular hits of the moment), the emergence of music videos and MTV revolutionizing the whole industry (Straw, 1993), and the rise in popularity of electronic dance music epitomized by the synthesizer (Goodwin, 1988). All this coincides with the Second British Invasion in the early-to-mid 1980s (Puterbaugh, 1983). Named after the First British Invasion of the mid-1960s, with

the outbreak of Beatlemania in the United States, the Second British Invasion was characterized by a dominance of UK pop artists on the American charts led by acts such as Duran Duran or the Human League. Just as Trudgill (1983) argued that British pop music acquired some validity during the First British Invasion, the Second British Invasion consolidated this legitimacy, which in turn likely impacted style-shift (linguistic and otherwise) be it away from or toward an American model.

A chronology of important events in popular music, which will be the basis for discussion in Part II is provided below. It bears mentioning that this timeline is not intended to be exhaustive. Rather, this aims to provide a general overview of some of the key events, milestones, and artists relevant to this thesis.

CHRONOLOGY

- **1920s** – Jazz era and 'birth' of popular music
- **1950s** – Popularization of rock and roll
- **1960s**
  - **1960** – The Beatles form in Liverpool
  - **1964** – Start of the First British Invasion
  - **1969** – Black Sabbath form in Birmingham
    – Judas Priest form in Birmingham
- **1970s** – Popularization of FM radio stations
  - **1975** – Iron Maiden form in London
  - **1977** – Def Leppard form in Sheffield
    – Ratt form in Los Angeles
    – Billy Joel releases best-selling album *The Stranger*
  - **1978** – Duran Duran form in Birmingham
  - **1979** – Approximate start of the New Wave of British Heavy Metal (NWOBHM)
- **1980s**
  - **1980** – Bruce Springsteen releases best-selling album *The River*
  - **1981** – Mötley Crüe form in Los Angeles
    – Phil Collins starts his solo career
    – Creation of MTV and popularization of the music video format
  - **1982** – Start of the Second British Invasion
    – CDs are introduced

- **1983** – Radio stations start appealing to more people with the Top 40 format (Danesi, 2007)
  – *Acts of Conflicting Identity: The Sociolinguistics of British Pop-song Pronunciation* by Trudgill (1983)
- **1984** – Generic stabilization and homogenization of Anglo-American pop music
- **1985** – Decline of the Second British Invasion

2.4 Summary and conclusion

This chapter explored the sociophonetics of popular music singing, focusing on style-shift as a significant aspect of linguistic variation in this context. Beginning with Trudgill's foundational study on British pop-song pronunciation, the chapter looked at the evolution of singing style in popular music, identifying the tension between adopting American-like pronunciation features and emphasizing one's local origins, from deliberate 'Americanization' to the enregisterment of Pop Song English (PSE) as a globally recognized standard without a specific American English association. Studies such as Beal (2009), Coupland (2011), Morrissey (2008), and Simpson (1999) expanded on Trudgill's findings, introducing new models and highlighting cultural, phonological, identity, and performance-based factors influencing pronunciation in music. The theoretical frameworks of style-shift explored in Chapter 1 were reintroduced and directly linked to the specific case of singing, relating the case studies reviewed to their conceptual counterparts. Finally, I attempted to approach the sociophonetics of singing voice pronunciation through a chronological lens, describing the process of enregisterment of a standardized singing style and integrating it into a broader timeline of popular music evolution. This chapter provides background for further exploration into the historical and cultural dimensions of style-shifting in popular music, and its implications for identity construction and genre conventions. However, these stylistic shifts are not solely driven by cultural or social factors. Speech and singing are different acts, and the latter's complexity cannot be ignored. As the concept of "singability" suggests (Morrissey, 2008), certain sounds are easier to sing than others, and the physical demands of singing undoubtedly influence how artists adjust their pronunciation. This highlights the need to examine articulatory considerations–in other words, how the mechanics of the singing voice may shape pronunciation choices. By exploring this connection in the following chapter, we can better understand how the specificities of the singing voice work alongside sociocultural factors in popular music pronunciation.

3 The Singing Voice: an Articulatory Perspective

Singing and speech have profound differences at the articulatory and phonetic levels. This chapter seeks to address the specificity of the singing voice in detail, exploring its distinctive characteristics and the articulatory frameworks that govern its production. Much of the existing research on the singing voice concentrates on its acoustic properties[1]–which will be covered briefly, though the focus of this chapter is on the articulatory aspects that define the singing voice and set it apart from spoken language. Section 3.1 will review the theoretical framework of articulatory considerations that will be applied to the singing voice, namely the principle of economy of speech gesture as described by Lindblom (1983). As this framework is primarily based on speech, Section 3.2 will turn to the specificity of the singing voice, looking at its differences from speech as well as providing a classification of different singing styles and vocal techniques. Lastly, the articulatory framework described in the first part will be applied to the singing voice, considering how notions of economy of vocal effort and 'singability' (Morrissey, 2008) may influence pronunciation choices. The aim of this chapter is therefore to establish singing as a specific object of study, and to open up the traditional sociolinguistic standpoint on pop song pronunciation toward an articulatory perspective.

[1] For a systematic review on analyzing and understanding the singing voice from a largely acoustic perspective, see Kob et al. (2011).

3.1 Articulatory framework

3.1.1 H&H Theory

Looking at the question of phonetic variation from a listener-oriented production perspective, Lindblom's 'Hyper-' and 'Hypo-' articulation Theory (H&H) (1990) described speech as adaptive. Speakers are influenced by competing demands, placing them on a spectrum tending toward either listener-oriented hyperarticulated speech or speaker-oriented hypoarticulated speech. This claim is based on observations concerning perception and production. Perceptually, listeners derive meaning from the signal through the discrimination of specific sounds. This is further facilitated by signal-complementary processes, such as frequency effects, which help the listener disambiguate information that may be incomplete or unclear in the signal. In terms of production, Lindblom underlines the fact that speech is a motor skill, an ability involving the movement and coordination of the muscles in the body to perform specific actions. As a motor skill like any other, speech often defaults to a low-effort form when there are no active constraints preventing this from happening[2]. In other words, articulatory effort tends to be minimized whenever possible in speech, following a principle of economy of speech gestures. In addition to this principle, plasticity is also an important characteristic of speech motor control. If articulatory effort in speech is minimized when unconstrained, when a difficulty arises such as the presence of a bite-block in a subject's mouth, the system will adapt and compensate to maintain a signal that is as authentic as possible.

Based on these observations, Lindblom posited that a speaker's speech will vary along a continuum of hyper and hypospeech, continually adapting to the demands of a specific situation. In other words, a speaker's goal is to produce utterances that are intelligible to the listener, while spending as little energy as possible in doing so. The metaphor of a tug-of-war between speaker desires and listener demands has often been used to illustrate the H&H Theory, as represented in Figure 3.1.

A one-dimensional continuum between speaker and listener has been recognized as an oversimplification since its very conception, and its author expected it to be revised. Aylett and Turk (2004) refined the H&H Theory in developing the Smooth Signal Redundancy Hypothesis (SSRH), proposing a model of the complementary relationship between acoustic and lexical cues. They still defend the image of a 'tug-of-war' drag-

[2]There has been debate as to whether notions of economy or efficiency can be applied to any and all motor skills. However, most researchers in this field agree that the economy principle applies at the very least to well-rehearsed, repeated, and goal-oriented motor behaviors (Sparrow & Newell, 1998), which speech is undoubtedly part of.

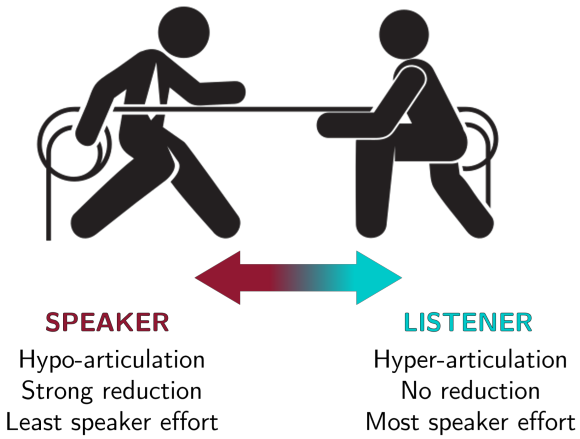


Figure 3.1: *An illustration of the H&H Theory by Lindblom (1990), inspired by a figure in Niebuhr (2016).*

ging speech along a continuum between two opposing forces, identified as *conservation of effort* and *reliable communication*. Specifically, they advance that the likelihood of recognition of an acoustic signal is directly tied to Signal Redundancy. Signal redundancy is a concept which combines both language redundancy–the likelihood of recognition based on linguistic elements such as the lexicon, syntax, semantics, pragmatics, etc.–and acoustic redundancy–the likelihood of recognition based on acoustic salience, such as distinctiveness, duration or loudness. According to the SSRH, when acoustic redundancy is low due to a lack of acoustic salience in the message, language redundancy will help facilitate intelligibility by compensating for the poor signal. Conversely, if language redundancy is low because a less frequent or predictable vocabulary is used, speakers will tend to increase acoustic redundancy (salience) to produce a balanced utterance. This theory places intelligibility at the center of a speaker's listener-oriented goals crucial to understanding variation in speech. It is important that intelligibility is taken into account in singing voice pronunciation, as singing is known to reduce it (Collister & Huron, 2008). This undoubtedly is a concern for any artist trying to make themselves understood, and likely an additional factor shaping their productions. This aspect, however, falls beyond the scope of the present study[3], which is primarily concerned with socioindexical and articulatory mechanisms rather than perceptual outcomes, despite the fact that the latter may influence the former.

3.1.2 Economy of speech gestures

One characteristic of speech motor control evoked in the H&H Theory and highly relevant to this thesis is the principle of economy of speech gestures, which can be defined

[3]A small-scale analysis of intelligibility factors can be found in Caillol and Ferragne (2024).

as the fact that "unconstrained, a motor system tends to default to a low-cost form of behavior" (Lindblom, 1990, p. 413). In other words, there is, in speech, a propensity toward under-exploitation, with the active goal of reducing energy expenditure as much as possible, as long as other demands (being understood by the listener, for example), are not negatively impacted. This is what is referred to as hypoarticulation.

Lindblom (1983) gives two examples of this principle of phonetic motor control in speech: consonant-vowel coarticulation and vowel reduction. Lindblom and Sundberg (1971), looking at specific tongue parameters in the production of stops, used a numerical model in order to assess the natural degrees of freedom of the articulatory system, and how tongue configurations may be influenced by the following vowel sounds. The model shows that normal speech uses only a fraction of the available degrees of freedom for articulation, avoiding extreme parameter values. In terms of vowel reduction, it is noted based on a number of previous studies, that it is similar to coarticulation in that its degree can be independently controlled, and lesser displacements and velocities are preferred. In other words, the production system is ruled by the overarching principle that extremes are to be avoided: when other factors permit, everything tends toward making speech more economical. The following section will identify specific characteristics of the singing voice, allowing for the application of the concepts in the current section to this specific area of study.

3.2 The specificity of the singing voice

3.2.1 Singing versus speech

Singing and speech share the same basic physical and physiological mechanisms, as shown in Figure 3.2. An over-pressure of air is created in the lungs via the act of breathing, which is then modulated into sound by the vocal folds through the phonation process. This sound then passes through the vocal tract, where its acoustic signal is shaped by the articulators such as the lips, tongue and jaw.

This is where the main similarities between singing and speech end, however. Singing constitutes a special case of voice use that requires an increased level of precision and coordination, with singers submitting their voice to conscious control producing the desired aesthetic effect of musical expression (Sundberg, 1990). To do this, singers must learn to systematically manipulate vocal fold vibrations and vocal tract configurations to create the desired sound with minimal effort while maintaining independent control over various vocal and articulatory parameters (Kayes, 2019; Sundberg, 1987, 1990). As such, the act of singing requires singers "to establish a

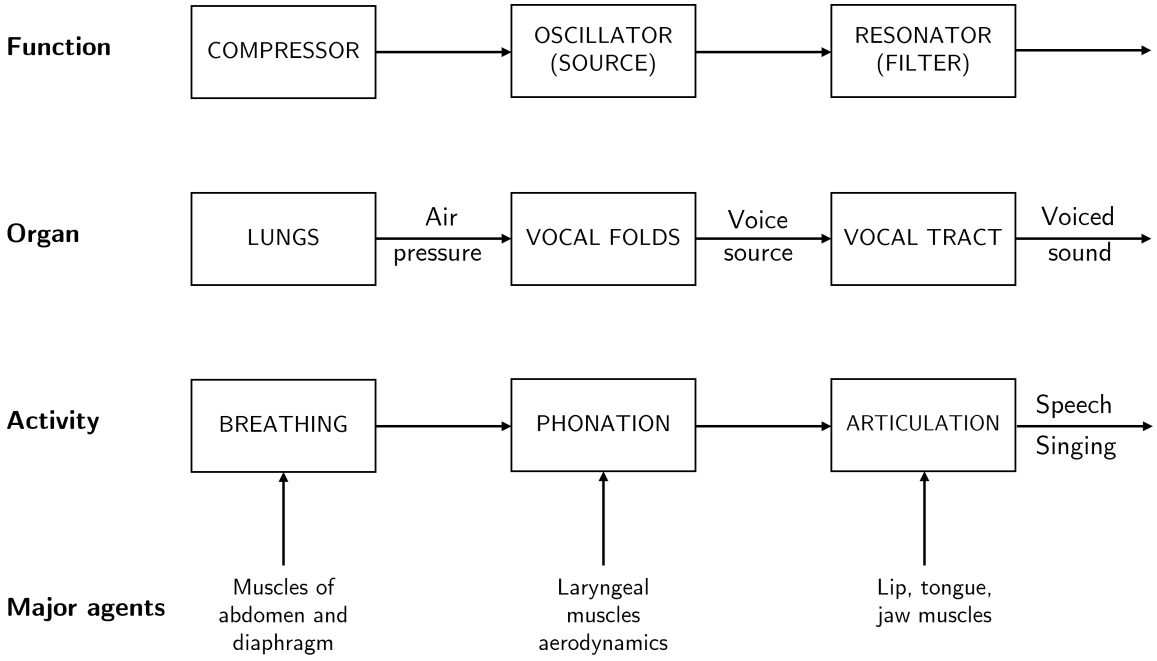


Figure 3.2: *The three functional constituents of the voice organ according to Sundberg (1987, p. 10).*

number of new patterns for breathing, articulation, and phonation" (Sundberg, 1990, p. 117). For instance, in speech, loudness and pitch are interdependent, and both ruled by subglottal pressure. When speaking, an increase in subglottal pressure leads to an increase in loudness, which in turn almost automatically leads to an increase in fundamental frequency. In singing, these parameters have to be independent of one another or pitch mistakes are sure to occur whenever loudness has to be modulated (Sundberg, 1990). Another parameter that differs from speech is the mode of phonation. One study comparing trained singers and non-singers found that singers were better able to maintain a specific mode of phonation when changing loudness than non-singers (Sundberg & Rothenberg, 1986). This further highlights how singing requires more independent control over phonatory parameters than speech. A further distinction between speech and singing concerns the relative duration of phonemes. In speech, the duration of syllables is usually ruled by word choice and idiosyncratic parameters. In singing, by contrast, syllable duration is determined by the rhythmic structure of the music. The syllable is significantly lengthened, which leads to an increase in the mean duration of vowels and a decrease in the mean duration of consonants. This is due to the fact that vowels are better suited to sustain the sound and melody of the song whereas consonants interrupt the musical line by blocking the flow of air (Scotto Di Carlo, 2007b). Figure 3.3, created for this work using data from Sergeant (2019), illustrates the relative duration of phonation (in percentages) of voiced versus voiceless

and consonant versus vowel sounds of fifty trained choristers speaking and singing a phrase from a traditional Christmas carol. Although the data is collected from children who are mostly used to singing in a choir, the proportion difference between speech and singing is quite striking. The proportion of phonation time of vowels in song as opposed to speech is greatly increased (from 59% to 90%) and that of consonants drastically decreases from 41% to only 10%. This figure is further discussed in Section 3.3.2, in relation to the concept of sonority.

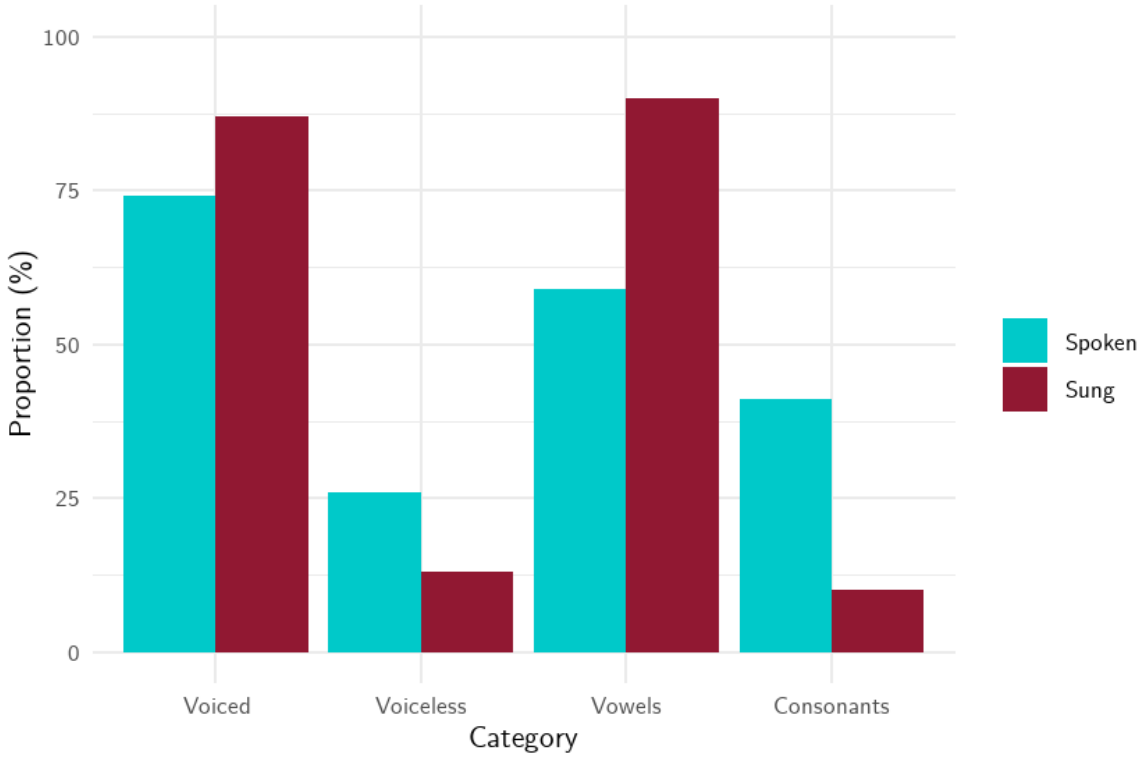


Figure 3.3: *Proportions of phonation time occupied by voiced vs. voiceless and vowel vs. consonant sounds in speech and singing, according to Sergeant (2019).*

The habits and patterns of singers and non-singers makes apparent the difference between speaking and singing, but the act of singing itself is far from being an homogeneous practice, and important variations exist between different singing styles. This thesis investigates heavy metal and pop music, which are presented along with different styles in the next section.

3.2.2 Styles of singing

From the Karnatic tradition of classical singing in India to the use of overtones in Mongolian Hoomii ("throat singing"), singing can take many different forms within a diversity of practices, techniques and functions. Following the area of study for this

thesis, this section will focus only on Western styles of singing[4], as this work itself examines only a very restricted aspect of vocal use within this category, namely pop and heavy metal singing.

Classical Western operatic singing

Classical Western operatic singing, also called Western lyric or more generally classical singing, is perhaps the most studied form of singing in all scientific areas concerned with vocal use. For a long time, the Western classical tradition has held an authenticity, legitimacy and authority that other singing styles have not (Potter, 1998). This status relies in part on the level of technicity required for classical singers, ruled by strict norms that take years to master. For example, unlike most contemporary commercial singers, classical singers do not (or rarely do[5]) use electronic amplification, necessitating alternate vocal solutions in order to be heard above loud orchestral accompaniment. As such, most male classical singers have developed a strategy known as the singer's (or singing) formant (Hollien, 1983; Sundberg, 1974, 1995). The singer's formant refers to a cluster of harmonic frequencies in the range of 3000 Hz, achieved by lowering the larynx. This creates a strong resonance, which allows the singer to be heard over the orchestra without the need for amplification and with minimal vocal effort once the technique is mastered. Another specific demand of classical singing, especially at high pitches characteristic of soprano singers, is formant tuning. In most cases, the fundamental frequency is lower than the first formant. In the upper parts of soprano singers' range (usually going as high as C6, or about 1445 Hz), however, this may no longer be the case, which leads to loss of vocal power and discontinuities in the musical line. Sopranos therefore deliberately 'tune' the first formant value to a value close to that of the fundamental frequency. This is mainly done by lowering the jaw or retracting the lips in a sort of smile to increase mouth opening, and thereby increasing loudness, uniformity of tone and ease of singing (Joliveau et al., 2004; Sundberg, 1987).

Contemporary commercial music

Most of what is known about the singing voice stems from studies on classical singers, who represent only a small and highly-trained group of professionals. Bringing more

[4]It must be noted that in this chapter, the term "style" refers to the classification of different forms or genres of singing, as opposed to the specific sociolinguistic concept of *style* and *style-shift*, defined in Section 1.2.1.

[5]There are of course cases of classical singers using amplification, especially in unconventional spaces or big stadiums, but this comes usually as a crutch. Though microphones are becoming more commonly used, they remain an exceptional choice in opera singing and this work therefore assumes that it has not, as of yet, irreversibly changed the way classical singing is taught and produced.

widespread types of singing into consideration may broaden our understanding of the singing voice. As Section 1.3.2 on defining popular music from a cultural studies perspective already hinted at, neatly categorizing different music genres based on specific characteristics is not as straightforward as it may seem. Contemporary commerical music (CCM) is a term coined by experts in vocal pedagogy to refer to a wide array of non-classical singing types (LoVetri & Means Weekly, 2003). It corresponds to a variety of genres, including but not restricted to popular music genres. Musical theater, Broadway singing or jazz are also represented. An important thing to note when it comes to CCM is that genre categorization usually concerns the artists themselves, as opposed to their actual productions. As such, though an artist may be identified as belonging to a certain genre, it is not unheard of that they may release music categorized in a different genre–pop music icon Beyoncé releasing a country album, *Cowboy Carter*, in 2024 being a recent example.

Though the borders between genres and subgenres may be blurry, one of the main characteristics of all CCM[6] is that, unlike classical music, singers usually resort to methods of amplification such as microphones in order to be heard (and, incidentally, to add real-time vocal effects). As such, certain techniques favored in classical singing do not need to be applied in modern singing, and most singers of CCM do not have a singer's formant (Burns, 1986). As Sundberg puts it, "we would not expect to meet the opera singer's technique in rock concerts, where sound amplification systems handle the balance between voice and accompanying instruments" (Sundberg, 1987, p. 124). Additionally, CCM singers do not necessarily have professional training (a necessity for classical singers), or may seek it later on in their careers, even if they are already successful (Fisher et al., 2019). Different genres may also pursue different aesthetic goals: various types of voice qualities or vocal techniques (nasality, breathy voice, creaky voice, grunting, belting, etc.) may be sought for different effects (Caffier et al., 2018), and a singer's individual touch can also matter greatly (Fisher et al., 2019). For instance, the use of vocal growl, characterized by simultaneous vibrations of the vocal folds and the upper parts of the larynx, is not something frequently heard in traditional classical singing, but is a well-used technique in styles ranging from *umngqokolo* (a vocal tradition of the Xhosa people in South Africa) (Sakakibara et al., 2004), to extreme metal music such as death metal (Guzman et al., 2019; Smialek et al., 2012). The introduction of these techniques in western popular music led to the observation that "just about the time that the recording industry has been able to remove all equipment noise and distortion ... performers are putting noise and distortion back into the sound

[6] CCM is a useful concept to define popular music in opposition to classical music and is used as such throughout this chapter, but popular music as defined in Chapter 2 remains the term of choice to refer to the genres studied in this work.

with their own voices." (Titze, 1997, p. 41).

Though it has not been studied as thoroughly as the singing voice in classical music, a few studies have looked at various characteristics of some CCM genres. Björkner (2008) and R. E. Stone et al. (2003) studied musical theater and Broadway singing, respectively, genres usually regarded as a hybrid between contemporary commercial music and classical operatic singing. Country singing has also been the subject of a series of studies (Burns, 1986; Hoit et al., 1996; Sundberg et al., 1999). More relevant to this research, as they are the two main genres under examination, are studies focusing on pop (including rock singing) and heavy metal. When compared to other singing styles and their phonetics features, pop and rock singing are fairly similar. Singers in both genres tend to show high F1 values and a 'speech-like,' neutral phonation, as well as having a high pitch range–though rock usually has a more narrow range than pop, allowing for singers to preserve its characteristic high energy (Thalen & Sundberg, 2001; Zangger-Borch & Sundberg, 2011). The main properties of heavy metal singing include pharyngeal constriction, raised larynx, tense vocal tract, complex modes of phonation (such as falsetto or creaky voice) and wide vocal range (Meireles, 2016). Musicologist Robert Walser (2015, p. 133) defines heavy metal singing as follows:

> The vocal sounds of heavy metal are similar, in some ways, to the guitar sounds. Quite often, vocalists distort their voices, for many of the same reasons that guitar players distort theirs. Heavy metal vocalists project brightness and power by overdriving their voices (or by seeming to), and they also sing long sustained notes to suggest intensity and power; sometimes heavy vibrato is used for further intensification (Rob Halford of Judas Priest is a prominent example).

Or, in the words of Halford himself, "metal is power." Very high pitched passages are also a feature of traditional heavy metal, and singers often use falsetto as part of their vocal technique in such cases, defined as the register used "when males attempt to imitate the female voice character" (Sundberg, 1987, p. 50). Guzman et al. (2014) argue that 'reinforced' falsetto (as opposed to naive or non-trained falsetto) is used in heavy metal, and associate it with a greater degree of vocal fold adduction, a shortening of the vocal tract and a very open mouth configuration. Another technique commonly used in heavy metal when singing higher notes is that of belting. Belting, originally used in musical theater, is defined as a speech-like or shout-like voice production, as a means of singing loudly and reaching high pitches (Popeil, 2007; Sundberg et al., 2012), leading professional singers of heavy metal to adopt open jaw and raised larynx settings (Meireles, 2016). Saturated voice or distortion is another a widespread supraglottal mechanism used by heavy metal and hard rock singers as timbral ornament. Described

in studies by Chevaillier et al. (2009) and Zangger-Borch et al. (2004), it is defined as a collection of 'dist-tones' all produced with different techniques, and deemed to be very taxing for the voice especially in highly demanding performances such as heavy metal concerts (Zangger-Borch et al., 2004). Differences of singing style concern all the main dimensions of phonation: fundamental frequency, loudness, voice quality, and formant frequencies, as well as articulatory settings and the techniques and expertise required to achieve a desired aesthetic goal. Though a majority of the literature describing the singing voice is concerned with classical singing, this thesis focuses on pop and heavy metal as over-arching styles.

3.3 Articulatory constraints applied to singing

The previous section described the specificities of the singing voice as opposed to speech, specifically within the wide range of different music genres and singing types. Singing techniques requiring specific practice and skill mostly concern classical singing (the singer's formant, formant tuning, etc.), but CCM singing as studied in this work also necessitates a certain vocal expertise that differs from regular speech. Additional articulatory constraints must therefore be considered, notably in terms of vocal economy and sonority.

3.3.1 Vocal economy in singing

If the principle of vocal economy outlined in the H&H Theory (Lindblom, 1990) applies to speech, it may very likely apply to singing too, as a type of vocal production that requires more focus, energy, and precision than regular speech (Sundberg, 1987). In a situation where the production system is pushed further than usual, the tendency toward minimizing vocal effort and optimizing energy expenditure will be greater. Singing imposes extra demands on the vocal system, such as maintaining precise pitches and often higher volumes, and the vocal system will try to find the most energy-efficient way to meet them. As a side note, live performance is also a major aspect of CCM (Bell & Gibson, 2011), and therefore includes modalities not directly linked to the singing voice but which may impact it nonetheless. For example, singers are expected to move (or even run) around the stage and dance, which adds extra physical constraints and thus further justifies the necessity to regulate energy expenditure. The ability to minimize effort while still being able to reach a particular aesthetic goal has been called "an essential aspect of the beauty of a voice," which suggests that "singers are more systematic in their voice use" (Sundberg, 1990, p. 117) than non-singers. In a questionnaire asking singers about their representations of vocal efficiency, a majority of

respondents insisted on the importance of production efficiency (Pillot-Loiseau, 2006). In other words, singing in a way which requires little effort.

There are two main ways in which this principle of vocal economy can be seen in the singing voice. The first concerns the pronunciation of vowels: "singers have a clear-cut tendency to centralize vowels in all registers" (Scotto Di Carlo, 2007a, p. 5). This can be explained by the need to keep the vocal tract unobstructed and lessen contraction in facial muscles in order to maintain a high degree of flexibility, necessary requirements for efficient singing. These adaptations inevitably lead to hypoarticulation, which, as the tug-of-war image illustrates (see Figure 3.1), comes at the price of intelligibility, especially at very high pitches. An example can be seen in Johansson et al. (1982), who investigated midsagittal contours of the tongue body for the vowels /ɑː/, /iː/, and /uː/ spoken then sung at three different fundamental frequencies (230, 465 and 940 Hz). Results showed that for the lowest pitch, the tongue positions for sung and spoken vowels were nearly identical. At 940 Hz, the same tongue configuration was used for all three vowels, which incidentally happened to be the tongue shape favored when producing /ɑː/, the most neutral vowel of the three. Lindblom (1990) cited this study as an illustration of his precept that motor systems tend to default to behaviors which prioritize minimal effort. While sustaining a vowel sound (especially a closed vowel) at a lower pitch does not appear to require more effort or energy than articulating it in speech, it is certainly much more challenging to do it at a higher pitch, which requires more overall tension of the vocal tract and a wider buccal opening. At higher pitches, the system compensates by resorting to a more neutral tongue configuration for all three vowels, thereby reducing strain and increasing loudness, albeit at the cost of decreasing the degree of differentiation between vowels.

The central role of vocal economy in singing can also be seen in the articulation of consonants. All consonants involve some degree of constriction or closure, which is incompatible with the necessity of having an open and unobstructed vocal tract when singing. The concept of *legato*, used in music theory and a common performance direction, is useful to understand the necessity of minimizing vocal effort in the singing voice. *Legato* is defined as the smooth and connected passage from one note to the other, without interruption of the musical line or abrupt changes in intensity. In other words, the transition from note to note should be seamless, without intervening silences or intermittent interference in the signal–which, by definition, consonants, and more specifically plosives such as /b/ or /t/, cause. The opposing concept of *staccato* describes musical phrases where each note is clearly designed with long pauses and typically a shorter duration. Similarly, fricative consonants such as /ʒ/ or /ʃ/ narrow the vocal tract and make audible friction noise that may disturb smooth *legato* phrasing

(Scotto Di Carlo, 2007a). Singers are thus prone to reduce consonantal duration (see for example Figure 3.3) and articulatory force, favoring vocal economy and hypoarticulation. Both vowel centralization and consonant reduction provide evidence of the tendency toward minimizing energy expenditure in contexts specific to singing.

3.3.2 A preference for sonority

Research into the sociophonetics of the singing voice (see Section 2.1) suggests that something like 'ease of singing' may be at play in addition to purely sociocultural arguments. Without making an explicit connection to Lindblom's theory, the few previous studies on this topic have approached it in terms of singability and sonority. Concluding his overview of different theoretical frameworks for pronunciation in pop singing, Trudgill stated that "even in the narrow field of pop-song pronunciation it is not possible ... to explain why particular ... consonantal, vocalic or other variants are retained, rejected or selected, and not others" (Trudgill, 1983, p. 159-160). With this remark Trudgill recognized that there must be one or more unknown variables influencing singing pronunciation, which cannot be explained by sociolinguistic theories, such as those described in Section 1.2.2 and Section 2.2.2. Although Trudgill called for some theoretical refinements to help complete the picture, this question has rarely been raised in the literature, Morrissey (2008) providing a noteworthy exception.

Since the data being studied in this domain does not come from speech but from singing, Morrissey (2008) argued that specific rules and restrictions apply and that traditional sociolinguistic considerations may have limited explanatory power. More specifically, he put forward the claim that "there are certain speech sounds that lend themselves to singing better than others" (Morrissey, 2008, p. 209), a concept he referred to as singability, and directly linked to the concept of sonority as singable sounds 'carry' the tune" [p. 210]. Sonority is defined as the carrying power of speech sounds, with more sonorous items having greater carrying power, and requiring less energy to produce acoustically (Hawkins, 1984). The hierarchy of sonority from most sonorous to least sonorous is shown in Figure 3.4. This suggests that more sonorous open vowels are preferred in singing to their least sonorous counterparts such as plosives or affricates, because they are better able to carry the melody. This can be seen in Figure 3.3 in Section 3.2.1, where the proportion of voiced sounds (87%) and vowels (90%) are markedly higher than voiceless sounds (13%) and consonants (10%). This indicates a clear preference for more sonorous sounds in singing. More generally, the claims of professionals involved in music (whether singers or phoneticians) align with the experiences of anyone who has ever tried singing can agree that it is difficult–and in some cases even impossible–to sustain a note on a consonant (see for example Scotto

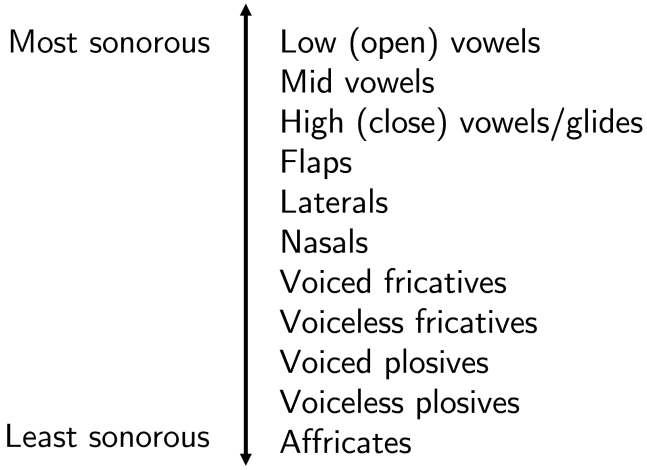


Figure 3.4: *The sonority scale according to Burquest and Payne (1993, p. 101).*

Di Carlo (2007b), Sergeant (2019), and Sundberg (1987)).

Morrissey (2008) argued that the features making up the USA-5 model (see Table 2.1) are typically more sonorous in American English than British English, which could explain why American English realizations are favored by singers. For instance, intervocalic /t/ is much higher up in the hierarchy of sonority when realized as a flap rather than a voiceless stop, with stops often more difficult to sing, depending on the musical phrasing of the song. The specific case of word-internal intervocalic /t/, the main focus of this thesis, will be considered in more detail in Chapter 4. Similarly, LOT unrounding, where the more open and less rounded [ɑ] is preferred over the more typically British [ɒ] provides another example of how sonority may affect pronunciation features. It is reasonable to conclude with Gibson (2023) that sonority is "an important non-social dimension likely to structure variation in song" (Gibson, 2023, p. 21). Sonority, and the related concept of singability strongly suggest that there are more than dialectal factors influencing singing voice pronunciation. Lindblom's concept of vocal economy (1983, 1990) can be applied in this context: the more sonorous allophones of American English are preferred because they require less energy and hold the melody better–put differently, they are easier to sing than their less sonorous British English counterparts. This tendency is especially pronounced in singers' attempts to reduce the articulatory force of consonants. The concepts of singability and sonority offer evidence of how phonetic choices may preserve vocal economy in singing.

3.4 Summary and conclusion

Moving beyond established sociophonetic theories, this chapter introduced an articulatory approach to the singing voice and the question of its influence on pronunciation. The first part of the chapter focused on building the theoretical framework for the notion of vocal economy, with Lindblom's H&H Theory underlining the importance of speaker-oriented considerations, ruled by the tendency to minimize vocal effort and reduce energy expenditure. This chapter firmly established its relevance to singing as well as speech, and identified contemporary commercial music (CCM), and specifically heavy metal and pop as genres where a lesser role of formal instruction than in classical singing may strengthen effects on pronunciation. Vocal economy is crucial for understanding the singing voice as singers experience more extreme constraints on the vocal system than when they speak. Although they share fundamental physiological mechanisms, speech and singing are inherently different cases of vocal use. Singing requires greater precision and coordination to achieve desired aesthetic effects as shown in Section 3.2 on the specificity of the singing voice, first compared to speech, then between different singing styles. Lastly, this chapter linked articulatory constraints to the specificity of the singing voice, examined how vocal economy applies to the production of vowels and consonants, and referred to examples in the sociophonetics of the singing voice literature which have hinted at this principle in the concept of singability.

Overall, this chapter demonstrated that the differences between British singers' spoken and sung pronunciation appear to be influenced not only by the sociocultural and dialectal factors raised in Chapter 2, but also by inherent characteristics of the singing voice. Vocal economy and the preference for sonority suggests that certain pronunciation choices may be driven by the need to minimize vocal effort in singing, perhaps even more so than in speech as the phonatory system is pushed to more extremes. Features in the USA-5 model (see Table 2.1) were identified as pronunciation traits where vocal economy could favor their American realization by British singers. The following chapter, Chapter 4, will focus on such an American pronunciation feature, the flap realization of word-internal intervocalic /t/.

4 The Case of Intervocalic /t/

Referring back to Chapter 1, where the concept of the sociolinguistic variable was defined, Labov (1966) used the various realizations of post-vocalic /r/ as an indicator of non-linguistic dimensions of the social context in which it is produced. But, more than just a group signifier, the sociolinguistic variable is a means for speakers to actively create and project a certain identity, and position themselves within or against a certain group at a given time (Eckert, 2012), an idea comprised within the concept of indexicality (Silverstein, 2003). The USA-5 model (Simpson, 1999) as outlined in Table 2.1 lists pronunciation tendencies in the singing of British pop music artists. It can also be seen as a collection of sociolinguistic variables indexing the identification of British artists to a specific geographical area and its population (the United States and American people) at a given point in time, one which has then evolved to become associated with pop music in general (see Section 2.3.1 for a more in-depth description).

This work focuses primarily on the first feature of the model, namely the pronunciation of intervocalic /t/, which, in words like *better* or *city*, tends to be pronounced by British singers as [d], rather than [t] or [ʔ] (Simpson, 1999), or, as Trudgill (1983) put it, "a voiced alveolar flap of some kind" (p. 141). Chapter offered the famous example of Paul McCartney in 'Hey Jude.' This chapter gives an account of the pronunciation of intervocalic /t/ in English, starting with an articulatory and acoustic description of its realizations, followed by a closer inspection of the flapped allophone specifically. The sociophonetic aspects of intervocalic /t/ will be discussed as a main feature of differentiation between American and British varieties of English. Lastly, this chapter will address the specific case of intervocalic /t/ in the singing voice, with regards to the articulatory elements of singing developed in Chapter 3.

4.1 The pronunciation of intervocalic /t/ in English

As its name suggests, intervocalic /t/ is a phoneme that occurs between two vowel sounds, either word-internally, such as in *city* or *butter*, or across word boundaries, as in *let alone* or *get away*. By definition, intervocalic /t/ cannot be in initial position. Sonorants that occur between the preceding vowel and /t/ (as in *winter*), or cases where /t/ is followed by a syllabic /l/ (as in *little*), are often also treated as intervocalic contexts. For the purpose of this work, the following discussions will only address the word-internal context (for more on this choice, see Section 4.2), starting with an overview of the different allophones of intervocalic /t/.

4.1.1 Allophones of intervocalic /t/

Stop /t/

The most canonical realization of intervocalic /t/ is likely the voiceless alveolar plosive [t], characterized by a clearly defined closure stopping the airflow, followed by a sharp release. Acoustically, its mean duration measured from F1 onset to release is around 130 ms, and the burst frequency is usually above 3 500 Hz (Klatt, 1975; Steriade, 2000; Zue & Laferriere, 1979). Figure 4.1 shows the waveform and spectrogram of a prototypical intervocalic stop /t/.

In terms of articulation, plosives such as [t] represent the narrowest degree of consonant constriction, creating a complete closure, tight enough that no air exits the mouth. Because this entails raising air pressure within the oral cavity, the tongue must not merely make contact, but compress against the alveolar ridge to create a tight seal. The requirement to apply substantial lingual pressure against the palate can lead to overshoot[1], where the physical target of the movement is actually located beyond the place of constriction, "as if in an attempt to penetrate the opposite surface" (Gick et al., 2013, p. 174). Figure 4.2 shows a right midsagittal view of overshoot in the vocal tract during the production of stop [t]. The hypothetical tongue target beyond the actual place of constriction is shown in dashes, illustrating the tight compression that occurs as the tongue presses against the alveolar ridge to the point of deformation.

A second articulatory characteristic of the realization of stop /t/ is the contact that occurs between the lateral borders of the tongue and the hard palate, also known as lateral tongue bracing. In other words, lateral tongue bracing refers to an intentional tongue posture whereby the sides of the tongue make contact with the sides of the

[1] This type of overshoot is not to be confused with what Bell and Gibson (2011) refer to as *stylistic* overshoot, a sociophonetic process which can be equated to hypercorrection.

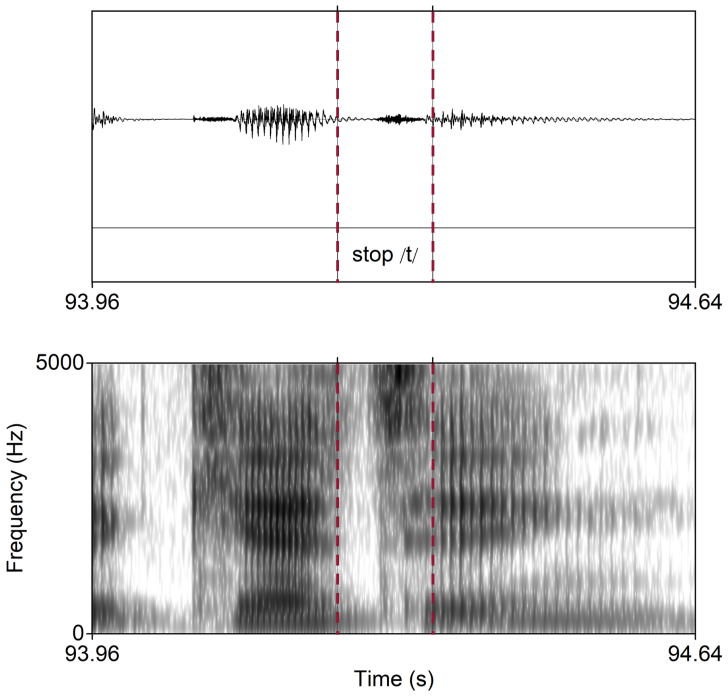


Figure 4.1: *Waveform and spectrogram of a prototypical stop /t/ realized by a native British English speaker. Image created in Praat (Boersma & Weenink, 2025).*

palate and the upper molars (Liu et al., 2023). The resulting tongue shape has been described as a "spoon-shaped" (Lee et al., 2015) or "horseshoe"-like configuration (Gibbon et al., 2007). Studies such as Lee et al. (2015) have found that a rapid rising of the lateral regions of the tongue occurs from the back toward the front in the millisecond before full closure is attained. In the production of alveolar stops, lateral tongue bracing plays an important role in controlling airflow direction, building up oral pressure, and providing overall tongue stability and somatosensory feedback to the speaker (Fuchs et al., 2006; Gick et al., 2013; M. Stone et al., 1992). The significance of this articulatory characteristic in intervocalic stop /t/ production will be further developed in Section 10.1.3, as it underlies the choice of electropalatography (EPG) as a technique to explore its realization in the singing voice.

Other forms of stop intervocalic /t/ exist in addition to the canonical alveolar plosive, including the glottal stop, where it is pronounced as [ʔ], with a closure of the space between the vocal folds (the glottis) followed by an explosive opening. An unreleased stop can also be found in intervocalic position, where /t/ is pronounced without the final release of air, as [t̚]. Other allophones of stop /t/ also include lateral and nasal release, [tˡ] and [tⁿ], respectively.

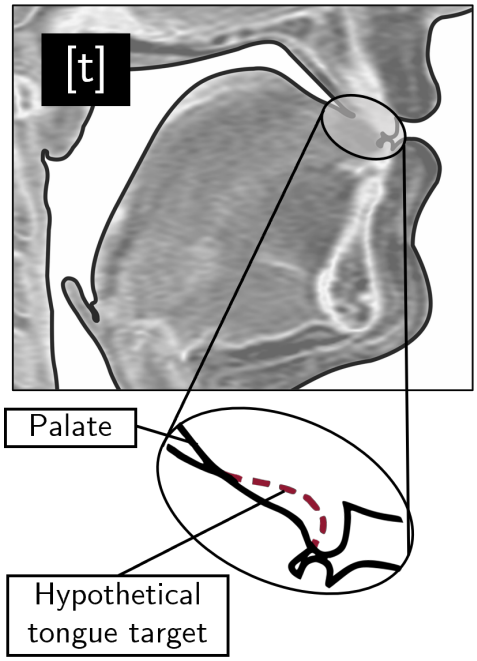


Figure 4.2: *Lingual overshoot in the production of [t]. Image by D. Derrick, figure inspired by Gick et al. (2013).*

T-Voicing

In some dialects including American English (see Section 4.2 below for more on the dialectal relevance of this feature), intervocalic /t/ can also be voiced and therefore tends to sound more like /d/ than in other accents. Though it does not necessarily imply a complete neutralization of the opposition between /t/ and /d/, T-Voicing may make words like *atom* and *Adam* or *writer* and *rider* sound homophonous.

The entirety of conditions ruling the production of voiced intervocalic /t/ in certain environments are not fully understood (Shockey, 2003), but experts tend to agree on the general rule first defined in the 1930s, whereby "it cannot occur unless it is *preceded* by a vowel or a sonorant (l, n, r), and is *followed* by an unstressed syllable-forming element (vowel, l, r, but not homorganic n)" (Haugen, 1938 as cited in de Jong (2011, p. 2), [original emphasis]). For the sake of simplicity, I will keep referring to /t/ in such places as 'intervocalic,' though as the definition suggests, this also includes some consonants. In simpler terms, a voiced /t/ can occur in both V_v and v_v environments.

Though the general rule applies in most situations where T-Voicing occurs, some specific cases may be less straightforward. Contexts such as /nt/ are one, where word-internal intervocalic /t/ is preceded by /n/. Vaux (2000) points out that sequences with this type of cluster are variable in their outcome, where they may either be realized with /t/ deletion followed by /n/ flapping in the case of *winter* or *hunter* or as a nasal

followed by a voiced flap (in *carpenter* or *seventy*). This may be explained partly by geographical variation, with speakers of Southern American English tending to simply delete /t/ altogether, as opposed to speakers of Northern American English who may preserve a clear distinction between *winter* and *winner* Wells (1982a). Overall, it seems that the conditioning of T-Voicing by a previous nasal appears more variable than other factors, and some speakers may resort to T-Voicing with nasals while others do not (de Jong, 2011; Zue & Laferriere, 1979). Other special cases have been identified, such as the fact that, according to the general rule stated above, the intervocalic /t/ in *Mediterranean* should be voiced (Steriade, 2000), or that a similar stress pattern in *capitalistic* and *militaristic* does not necessarily lead to the same realization, where /t/ is voiced in the former but not the latter word. Variability in the outcome of the T-Voicing rule is therefore a main component of the pronunciation of intervocalic /t/, and there are many exceptions to its systematic application.

When the general rule does apply and /t/ is voiced, by definition the vibrations of the vocal folds do not stop as they would for a voiceless [t], but continue throughout the alveolar articulation. As such, there is little or no buildup of air pressure at the place of articulation, and as a result, there is little or no stop burst at the closure release. It is furthermore characterized by its very short duration, with numbers varying from 26 ms on average (Zue & Laferriere, 1979) to a range of 20-40 ms for different speakers (Fukaya & Byrd, 2005; Herd et al., 2010; Shockey, 2003). /t/ has been referred to as "the shortest consonant in English" (Umeda, 1977, p. 849), where "[its] most salient acoustic feature is not ... its voicing–which need not even be present–but rather its shortness" (Trask (2001), as cited in Iverson and Ahn (2007)). Figure 4.3 shows the waveform and spectrogram of a prototypical intervocalic flap /t/.

T-Voicing discussed in this section should be understood as a phonologically-oriented, generic term that covers multiple phonetic realities. At this level of analysis, it can therefore simply be transcribed as a /t/ with the diacritic indicating voicedness: [t̬]. However, this thesis is concerned with more detailed articulatory descriptions, and from this perspective, T-Voicing can be seen as only an optional secondary step after a *flapped* or *tapped* articulation in the process of full neutralization of the /t/-/d/ contrast (Wells, 1982a). There is disagreement on the classification and transcription of this trait on an articulatory basis, alternatively (and sometimes interchangeably) referred to as tapping and/or flapping. Although these two specific phonetic realizations remain to be more precisely examined, the next subsection will outline the major differences between the two, and justify the preference for the term *flap* over *tap* for the rest of this work.

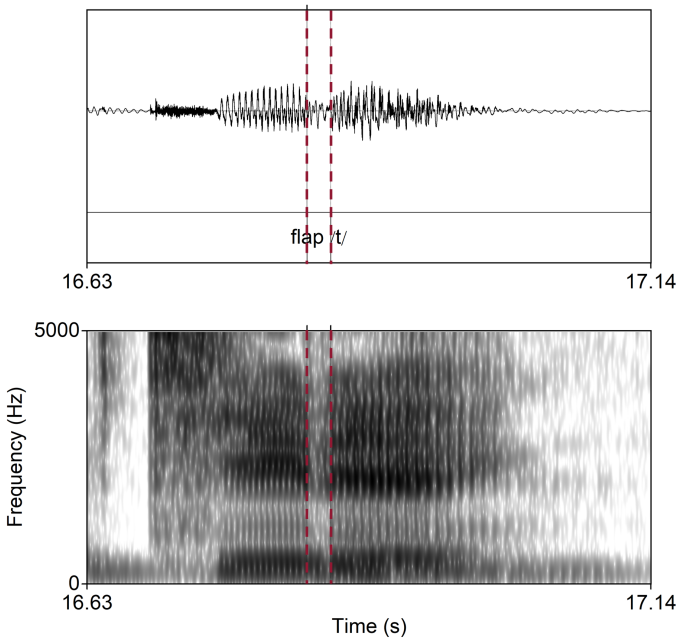


Figure 4.3: *Waveform and spectrogram of a prototypical flap /t/ realized by a native American English speaker. Image created in Praat (Boersma & Weenink, 2025).*

A terminological precision: flapping versus tapping

While some use the terms interchangeably, researchers agree that flapping and tapping refer to two different articulatory realities. The distinction between flaps and taps suggested in Ladefoged and Maddieson (1996) relates to the movement of the tongue tip when it makes contact with the alveolar ridge–in other words, the degree to which the motion of the tongue is tangential to the point of articulation. Taps are typically made by a rapid and direct movement of the tongue tip against the alveolar ridge, where the tongue begins and ends its movement in the same relative position. Gick et al. (2013) refer to it as a very quick or light stop, with very little to no overshoot at the constriction site. In the same vein, Laver (1994) and Wells (1982a) describe the tap as a limited example of a trill, with only one contact of complete closure. Flaps, on the other hand, are described as a *ballistic* movement (Wells, 1982a): it involves the tongue tip hitting the alveolar ridge in passing, and continuing along a single trajectory. Therefore, some overshoot does occur, but tangentially to the passive articulator. The tongue briefly touches the surface and then continues its movement in the same direction, either upward or downward (de Jong, 2011; Gick et al., 2013). As such, the tap primarily involves tongue movement orthogonal to the articulator, as opposed to the flap which tends to involve a large component of parallel movement.

Based on this difference, de Jong (2011) noted his impression that the /t/ in *city* is typically more of a tap, whereas the /t/ in the name *Berta* would more likely be a flap. Derrick and Gick (2011) took the distinction between flapping and tapping a step further, identifying four subphonemic kinematic variations in the pronunciation of a flapped and tapped /t/ using ultrasound data. Table 4.1, reproduced from their paper, provides a description of these variants (where TT refers to the tongue tip), and Figure 4.4 represents them as diagrams from a right midsagittal view of the vocal tract.

Symbol	Name	Description
[ɾ↕]	Alveolar tap	TT below the alveolar ridge, moves up to hit the ridge and back down below the ridge.
[ɾ↘]	Down-flap	TT above the alveolar ridge, moves down to hit the ridge and continues downward below the ridge.
[ɾ↖]	Up-flap	TT below the alveolar ridge, moves up to hit the ridge and continues upward above the ridge.
[ɾ↔]	Postalveolar tap	TT above the alveolar ridge, moves forward to hit the ridge and back again above the ridge.

Table 4.1: *Four flap/tap variants in English, according to Derrick and Gick (2011, p. 309).*

Though there is still ongoing debate over the terminology of tapping versus flapping and their status as inherently different realizations, they are ultimately not contrastive, at least in English. For the sake of clarity, I have elected to follow de Jong (1998, 2011)'s reasoning and refer to this phonetic realization as a flap rather than a tap. Even though phonetic classifications tend to make a proper distinction between flaps and taps, flapping is "the most typical, majority name for the phenomenon of interest" (de Jong, 2011, p. 3). Research on the topic tends to acknowledge that there is a difference between flaps and taps, but most choose to not make a terminological distinction between the two, usually using the term 'flap' as an umbrella term to refer to the general [ɾ] sound, whether or not the authors agree on the actual phonetic reality of the phenomenon (see, for example, Fukaya and Byrd (2005), Iverson and Ahn (2007), and Warner and Tucker (2017)). A production experiment conducted by Derrick and Gick (2011) has shown that, although English speakers who display this allophone produce both flap and tap variants, with a high degree of within and between-subject variability (participants produce different variants of intervocalic /t/ across repetitions of the same word in the same context), there is an absolute preference for flaps over taps. While the difference between flaps and taps is an important one, further work remains to be done on the articulatory and acoustic aspects of these variants in English

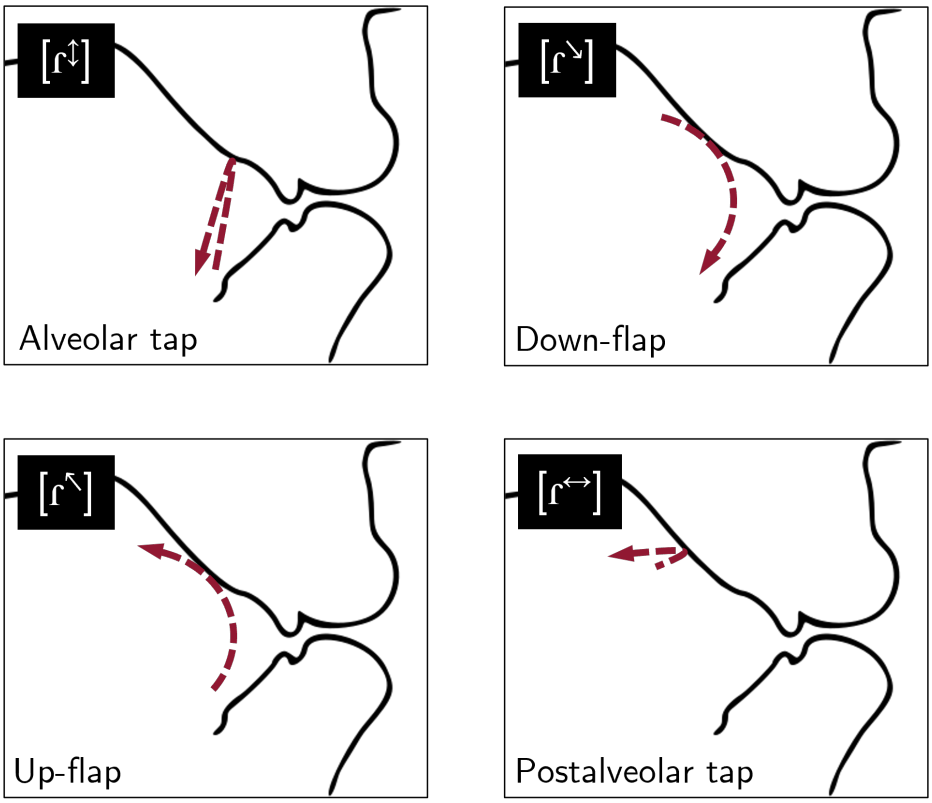


Figure 4.4: *Diagrams of the four flap/tap variants, according to Derrick and Gick (2011, p. 310).*

before a conclusion can be reached about the relevance of this terminological distinction (de Jong, 2011). Therefore, flapping in the rest of this work will refer to the articulation of intervocalic /t/ as [ɾ], regardless of whether the actual realization is a flap or a tap, as long as it is perceptively different from a stop /t/.

4.1.2 Flapping in focus

The acoustic characteristics of the word-internal intervocalic flap in English revolve around two main criteria evoked above: voicing and duration. Herd et al. (2010) confirmed the importance of duration to distinguish between flaps and stops, using speaker-by-speaker cutoff points: they found a mean of 56 ms as the average flap cut-off value, based on the gap between flaps and stops in speakers' individual /t/ distributions. Further acoustic correlates of flapping, including intensity, F4 measures, and effect of careful vs. spontaneous speech style have also been studied, notably in Warner and Tucker (2011, 2017) and Warner et al. (2009).

Looking more closely at the articulation of flaps, some research has attempted to show that flapping may arise naturally from the dynamics of the speech motor system, rather than from a distinct categorical speech goal. Fujimura (1986) proposed that flapping might result from interactions between tongue and jaw movements, which gives the feature its characteristic acoustic effects. After conducting articulatory analysis of flapping using X-ray data for tongue point trajectories in the sagittal plane (an articulation that is usually linked to the position of the tongue body and the jaw), de Jong (1998) posited that, according to jaw dynamics, a flap should result from lower jaw position and a tongue body position less likely to be a hindrance to the realization of an alveolar closure. No data confirming this was found, with no systematic or significant difference in jaw position during the articulation of flaps and stops which could have allowed for a clear articulatory explanation. Looking at kinematic data from the tongue tips collected with electromagnetic articulography (EMA), Fukaya and Byrd (2005) also attempted to explore the articulatory foundation of the distinction between flaps and stops, finding that flaps do not differ categorically or uniformly from stops in the articulatory domain, despite systematic shortness in terms of duration in the acoustic domain. The general pattern they were able to identify was that flaps are produced by reduced occlusion in the consonant. An implication of this finding is that the flap configuration is more similar to the configuration for vowels than for stops. Fukaya and Byrd (2005) concluded that while a biomechanical model cannot fully explain flapping, it does appear that the outcome of flapping is more optimized for speech movement. The exact causal relationship between acoustic output and articulatory dynamics remains unclear, but flapping seems to be influenced by both, and can be

placed along a continuum of weakening from a fully articulated [t] to a flap.

If flaps require less effort to produce, then the principle of economy of speech gestures should encourage its production. This idea can be explained as follows:

> If a simple articulatory gesture works just as well as a complex one, there is a natural tendency to prefer it, thus rendering the articulatory movements in speech simpler. For example, a voiceless [t] between vowels, as in *better* or *atom* involves not only a tongue-tip movement up to the alveolar ridge and away again, but also a switching off and on again of the vibration of the vocal cords. It is simpler, and requires fewer motor commands to the organs of speech, if the vibration of the vocal cords is continued throughout the alveolar articulation. (Wells, 1982a, p.94)

This is in direct relation to Lindblom (1990)'s H&H Theory as discussed in Section 3.1, where flap /t/ can be defined as a hypoarticulated variant of stop /t/. Westbury and Keating (1986) stated that they hypothetically agreed that a medial stop found between two vowel sounds should naturally be voiced, and is relatively easier to produce than other stops. Looking at world languages data, they did not find that speech gesture economy could be seen as the primary determinant for stop consonant voicing. In English, however, and specifically in American English (see Section 4.2 below), this type of reduction for intervocalic /t/ is well documented, and occurs to a greater extent in connected and spontaneous speech (Warner & Tucker, 2011; Wells, 1982a), though it has also been found to be a feature in careful speech (Shockey, 2003). Taking it one step further, Warner and Tucker (2011) found evidence of intervocalic flaps being realized as an approximant, or approximant-like token, highlighting the strong reduction this phoneme may undergo, to the point where the articulators are not making contact with one another. Though more research is required to confirm whether the English flap is 'easier' to produce, it does represent a weakened/reduced end of the continuum of possible intervocalic /t/ realizations. As previously mentioned, however, this seems to be the case for American English first and foremost, as flapping is one of the main characteristics of this specific variety of English.

4.2 Sociophonetic aspects of intervocalic /t/

Before discussing in more detail the sociophonetic aspects of the pronunciation of intervocalic /t/ to describe how it has been–and to a certain extent still is–a clear feature of American English (AmE) as opposed to British English (BrE), a caveat is in order regarding what is meant by these two varieties in this thesis. American English is

here understood in the way that Wells (1982a) defines General American: as a generic, unmarked accent that can be found in the United States and some parts of Canada. Though it may be somewhat questionable in terms of dialectal reality, this definition covers a broad range of realizations and is therefore convenient as an umbrella term. This thesis will rarely go into more specific dialect nuances than between American English and British English.

The latter, however, is slightly more complicated to define. For the purpose of this thesis, what I refer to as British English can more accurately be described as Anglo-English, as used in King and Ferragne (2020), as the "English spoken in England" (p. 1). That is to say that varieties of Scottish, Irish, and Welsh English which are commonly included in British English are not considered here, but what is being studied is broader than what traditional Received Pronunciation (RP), Standard Southern British English (SSBE) or BBC English labels might cover.

These terminological decisions do not have theoretical pretenses. Rather, they allow the making of meaningful comparisons within the geographical and sociocultural perspective of this thesis on music productions by 1980s artists. The choice to adopt a broad distinction between two general categories, American English and British English, was intentional. These categories do not reflect unified dialectal or geographical realities, but are used to include the wide variety of backgrounds represented by the artists under analysis. This decision also aligns with a more layman-oriented understanding of style shifting in popular music. It reflects how the phenomenon is likely perceived by those directly involved—namely, British artists from the 1980s and their audiences. These individuals may have a general sense of what constitutes an American versus a British accent, even if they cannot identify the finer nuances or specific regional features of each.

4.2.1 Flapping: a defining feature of American English?

Flapping in American English

A phonetic trait included in the USA-5 model, flapped intervocalic /t/ is a distinctive feature of an American English (AmE) accent in speech. Most works on flapping mention that it is a defining characteristic of an AmE accent, as it has been noted since the mid-1930s by scholars such as Kenyon (1936) and Haugen (1938, as cited in de Jong (2011)). Since then, it has become established as "one of the most striking characteristics of American pronunciation to the ears of a non-American" (Wells, 1982a, p. 248), and the general rule mentioned in Section 4.1.1 is almost always obligatory in American English (Shockey, 2003). In a large-scale study of conversational speech

using the SWITCHBOARD corpus (Godfrey et al., 1992) of telephone conversations recorded in the US, Patterson and Connine (2001) found that more than 93% of tokens (out of a little over 4 000 tokens total) with an intervocalic /t/ were realized with a flap. In a more controlled experimental setting, Zue and Laferriere (1979) found that 99% of expected flaps were realized as such. Accordingly, it appears clear that flapping is by far the dominant variant in the production of intervocalic /t/ in AmE, and therefore can be said to be one of its main differentiating features when compared to other varieties of English.

Flapping in British English

Though flapping is a defining feature of American pronunciation, it is not entirely absent from British English (BrE) either. Generally speaking, intervocalic /t/ tends to be realized as either an alveolar stop [t] or a glottal stop [ʔ] in BrE–the former being overall largely preferred (O'Connor, 1991). Nonetheless, Cockney English is a notable example of a British variety where flapping does occur quite frequently, though more so across word boundaries in commonly used lexical items (*lot of*, *but I*, etc.) (Tollfree, 1999; Wells, 1982b). Yet, while it is a possible allophone in Cockney, in environments where both flaps and glottal stops can be realized, Tollfree (1999) notes that glottalization is systematically favored over flapping. Nevertheless, researchers who were active around the period of this thesis' musical corpora production have predicted the expansion of this feature to all varieties of British English (O'Connor, 1991) and beyond, seeing flapping "as the first distinctively American phonetic innovation likely to spread in time to all accents of English" (Wells, 1982a, p. 250)

A recent study by Alderton (2022) looked into the use of flapping in young speakers of SSBE from high socio-economic backgrounds, finding that the flap variant is used to index both authority and informality. In other words, it is being used as a tool for 'elites' to convey both authority and legitimacy while presenting themselves in a more approachable and less snobbish fashion. Interestingly, he posits that this link may come from the association of flapping with American English and its use in popular media, echoing the discussion in Section 1.3.1 on the link and influence of media on language change. Despite some evidence of an increase in the realization of word-internal intervocalic /t/ as a flap in both read and conversational speech of young SSBE speakers, flapping is far from being used to the same extent as its two other main British English variants, as can be seen in the distribution pattern found by Alderton (2022) and summarized in Table 4.2. Looking at two phonological contexts where flapping is possible (word-medial tokens preceding a syllabic /l/ and word-medial pre-vocalic tokens), the overwhelming majority of tokens are realized as alveolar and

glottal stops (88% and 92%, respectively, compared to 10% and 7% of flaps). Therefore,

/t/ Realization	LITTLE	PRETTY
Alveolar stop [t]	47	80
Glottal stop [ʔ]	41	12
Alveolar flap [ɾ]	10	7
Other	2	1
Total	**100**	**100**

Table 4.2: *Distribution of /t/ realization in two phonological contexts (LITTLE = word-medial tokens preceding a syllabic /l/, and PRETTY = word-medial pre-vocalic token), according to Alderton (2022, p. 291)*

while flapping is undoubtedly used in British English, it cannot be said to be the overall preferred allophone of /t/, but rather an optional contemporary variant used for specific social-indexical purposes. As part of this thesis, an experiment on the evolution of the perception of flapping in American and British varieties of English was conducted and is detailed in Chapter 8.

Overall, if only in terms of distribution and frequency of realization in speech, flapping remains first and foremost a markedly AmE feature, especially in the word-internal position. Though flapping can occur across word boundaries, this specific environment was not studied in this work because it tends to be slightly more frequent in British English and is therefore a less effective marker of AmE. Even if word-internal flapping is spreading to other varieties of English, and specifically to BrE, the remainder of this work assumes it was not as pervasive in the time period under study, namely the 1980s. At the risk of reducing the complexity of real-world BrE, methodological constraints were necessary to keep the scope of analysis attainable in this thesis. Consequently, in word-internal intervocalic environments where a flap may occur, both alveolar and glottal stop /t/ are considered typically British realizations, whereas a flapped /t/ is associated with American English.

4.3 Intervocalic /t/ in singing

Section 3 of Chapter 3, which was dedicated to the singing voice, discussed singing from an articulatory perspective and explained how, since singing requires more focus and effort than speech, strategies may be adopted on the part of the singer to deal with these added difficulties. Having established a general framework for the principle of vocal economy and singability, this section seeks to explain how the articulatory features of intervocalic /t/ influence the way it is sung. In other words, by exploring

how voicing, duration, and articulatory features of stop and flap /t/ meet the demands of vocal performance, the aim is to understand why one realization (a flap) may be preferred in a sung context over another (a stop).

4.3.1 Features of stop /t/ vs. flap /t/ in singing

Physical properties

A well-formed, prototypical alveolar stop requires both overshoot and lateral tongue bracing (see Section 4.1.1) in order to create a tight enough seal to build the intraoral pressure needed for a proper burst release. This interrupts voicing and takes longer to produce than a flap, which does not require full closure. Though de Jong (1998) did not find direct correlation between jaw dynamics and a preference for flapping in speech, his results showed that flap production involves less occlusion. As such, flaps have a less precise articulatory trajectory and a lower degree of tongue constraint. This implies that less articulatory effort is required to produce flaps than to produce stops, and that flapping may better suit a wider buccal aperture as required for singing (Lindblom, 1990; Scotto Di Carlo, 2007a). This is something which cannot be neglected by singers looking to minimize vocal effort and optimize energy expenditure. Additionally, and directly resulting from the physical properties of stops and flaps, both voicing and duration must be considered in more detail to understand how they influence the sung realization of intervocalic /t/.

Voicing

A stop /t/ is voiceless by definition, leading to a known physiological difficulty in producing and maintaining voicing during a stop (Iverson & Ahn, 2007). A flap, on the other hand, is almost systematically voiced. Figure 4.5 represents data collected using electroglottography (EGG) from previous research on intervocalic /t/ (Caillol & Ferragne, 2024). EGG is a non-invasive technique used to measure the rate of vibration of the vocal folds, and can therefore be used to study voicing, where a rise in the resulting waveform indicates closing of the vocal folds and a dip their opening. The uniform, constant waveform on the right of the figure shows the flapped realization of the intervocalic /t/ in the sung word *little*, and implies consistent voicing, whereas the visible flattening of the curve seen on the left for stop /t/ shows that the vocal folds have stopped vibrating.

As discussed in Section 3.3.2, there is a distinct preference for sonority in singing. This means vowels, or at the very least more sonorous voiced consonants, are easier to sing since they allow for a continuous flow of sound. The more sonorous the sound, the

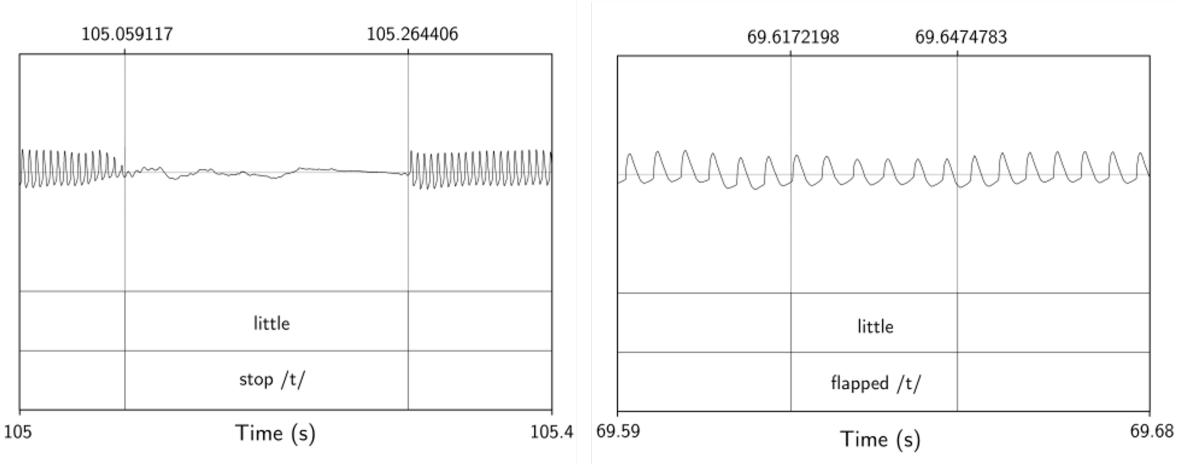


Figure 4.5: *Electroglottography (EGG) waveforms of stop /t/ (left) and flap /t/ (right) in the word* little.

better it is able to 'carry' the tune and it requires less energy to produce acoustically. Since flaps, but not stops, are voiced, flapping may be favored in singing as it will not hinder the vibration of the vocal folds. This characteristic allows the singer to maintain voicing throughout the musical line and keep *legato* phrasing undisturbed. This relates to the principle of least effort inasmuch as it can be applied to singing, as "it is simpler, and requires fewer motor commands to the organs of speech, if the vibration of the vocal cords is continued throughout the alveolar articulation" (Wells, 1982a, p. 94). Referring back to Figure 3.3, Sergeant (2019) found voiced sounds occupy more than 87% of phonation time as opposed to voiceless sounds when singing (against a proportion of 74% in speech), supporting the hypothesis of the least effort principle in singing.

Duration

Duration is another feature of intervocalic /t/ which influences the way it may be sung. Flaps are characterized by their voicedness, but also by their brevity in relation to stop /t/ in terms of closure duration (Iverson & Ahn, 2007; Steriade, 2000; Zue & Laferriere, 1979). Stops have been found to last around 100-150 ms, as opposed to 20-40 ms for flaps. As previously noted in Section 3.3.2, singing tends to reduce consonantal sounds, both in terms of duration and articulatory force. As such, a realization of /t/ which is inherently faster (a flap), may be preferred over one which takes significantly longer to produce. Attributes of flapped intervocalic /t/ (voicing and duration as a result of physical articulatory properties) can be taken as a probable argument as to why it may be favored in singing, considering articulatory constraints specific to the singing voice.

4.4 Summary and conclusion

This chapter focused on the main pronunciation trait that will be discussed in the remainder of this work, namely the distinctive realizations of word-internal intervocalic /t/ as an alveolar stop [t] and an alveolar flap [ɾ]. The first part described these allophones in acoustic and articulatory terms, before focusing on the specificities of the flap itself. The second section addressed the associated sociophonetic aspects of each realization, as features of British and American English, respectively. A few terminological choices regarding the use of terms such as flapping, British English and American English were explained, as well as the decision to restrict the scope of the analysis. Lastly, the features of the stop and flap realizations of intervocalic /t/ were reconsidered in the context of the singing voice and its inherent constraints, keeping in mind previously defined concepts of least effort and economy of speech gestures.

Examining the pronunciation of intervocalic /t/ as a sociolinguistic variable can highlight dialectal differences between American and British English. Not only can its pronunciation index geographical or group belonging, it may also reflect a more recent enregistered meaning for artists as a feature of Pop Song English (PSE), as discussed in Chapter 2. Exploring the articulatory aspects of the stop and flap allophones of intervocalic /t/ adds a further dimension that must be taken into consideration: one realization of the phoneme, namely the flap, may be easier to articulate in singing, considering factors like ease of pronunciation and minimal effort. This suggests that beyond the sociocultural use of /t/ as a sociolinguistic variable that indexes identity, its prevalence in the singing pronunciation of British artists–who do not typically exhibit this trait in their speech–may also be influenced by phonetic and articulatory factors inherent to the constraints of the singing voice.

Part I: Summary and Research Questions

4.5 Summary

Part I presented a general overview of the different fields and concepts which serve as the basis for this work and which will be referred to throughout. A concise summary of the main points of each chapter follows:

Chapter 1

This chapter highlights the importance of audience and speakers' agencies in the construction of their identities as the driving forces behind style-shift. In the Third Wave of sociolinguistic variation studies, speakers are no longer considered passive carriers of dialect features, but use such features in different ways and at different times in order to construct certain identities or position themselves within or against certain groups. Linguistic variation is deeply embedded in social meaning and identity construction, and popular music performances provide a unique study perspective of this phenomenon in the field of sociolinguistics.

Chapter 2

This chapter presents the methodological outlines of the sociophonetic approach to the singing voice taken in this work. A review of the literature on the subject considers the adoption or rejection of an American-influenced pronunciation style as a tool for identification, both perceived and projected. The theoretical evolution from 'Americanization' to Pop Song English follows historical changes of the music industry in the 1980s, and is highlighted as a turning point in popular music, that will be addressed throughout this work.

Chapter 3
Studying the singing voice requires considering it in relation to speech, with a focus on its distinguishing characteristics. Differences between British singers' spoken and sung pronunciation may be influenced not only by sociocultural and indexical factors, but also by inherent constraints proper to the singing voice. Notions of vocal economy and a preference for sonority suggest that certain pronunciation choices may be driven by the need to optimize vocal effort in singing, as the phonatory system is driven to more extreme conditions than in regular speech.

Chapter 4
Word-internal intervocalic /t/ as the linguistic variable of choice for this work offers an ideal focal point: it is specific enough to allow for detailed analysis, yet its relevance across multiple domains makes it a useful lens through which to approach the different aspects of this work. It represents both a marked feature of American-influenced pronunciation and/or PSE, and its variants are also interesting in terms of articulation, offering opportunities to examine how constraints specific to singing may have an impact on its realization.

The main motivation behind this thesis is to present an experimental approach to the existing body of literature on the sociophonetics of the singing voice in popular music. More specifically, it focuses on the well-documented tendency for British popular singers to use pronunciation traits resembling an American accent, through a larger-scale study of British heavy metal and pop than what has previously been conducted. It seeks to introduce a complementary perspective to the sociocultural argument traditionally put forward in terms of identity construction and genre belonging: an articulatory dimension which is still lacking or underdeveloped in most analyses of the phenomenon and that takes into account the particularities of singing, not least its added stress on the phonatory system.

4.6 Research questions

Part I provided a theoretical introduction to the sociocultural and articulatory phonetics of singing voice pronunciation, from which the following research questions that serve as the basis for this work emerge:

1. What can be learned from the different genres of the 1980s, as a decade of change in popular music, in terms of the tendency for British singers to style-shift?

   (a) Do genres with contrasting sound and image such as traditional heavy metal and pop music both show this tendency? To what extent?

   (b) How do their respective artists navigate the construction of their artistic identities and genre conventions?

   (c) In addition to intervocalic /t/, do other acoustic features also corroborate this tendency?

   (d) From a broader perspective, is the stop vs. flap allophonic distinction of intervocalic /t/ still as relevant today as it was in the 1980s?

2. Beyond sociolinguistics, and based on the phonetic features of intervocalic /t/ variants, how may articulatory constraints inherent to the singing voice contribute to this modification of pronunciation by British artists? In other words, can a flap /t/ or a flap-like realization occur naturally in singing as a result of the dynamics of the motor system looking to minimize effort, rather than as a distinct indexical goal?

   (a) To what extent is mouth opening an important factor in the difference between singing and speaking?

   (b) To what extent do certain factors of the music (pitch, syllable duration, etc.) dictate which variant of intervocalic /t/ is produced by contemporary British popular music singers?

   (c) How is the articulatory realization of sung intervocalic /t/ impacted in terms of linguopalatal contact and vocal tract configuration?

3. How do these articulatory and sociocultural constraints work together or against one another to shape what singers ultimately pronounce when singing?

Part II

Sociocultural factors

Part II: General Introduction

THIS SECOND PART constitutes the sociocultural and socioindexical component of my thesis. Positioned within the scope of variationist linguistics described in Chapter 1, and building on previous works on the sociophonetics of the singing voice as examined in Chapter 2, it examines specifically two genres from the 1980s: traditional British heavy metal and pop. In general, it seeks to answer the first research question posed in Section 4.6: What can be learned from the different genres of the 1980s, as a decade of change in popular music, in terms of the tendency for British singers to style-shift?–as well as its accompanying sub-questions.

This section explores sociocultural factors such as the dominance of the American music industry, the influence of musical role models, and the conflicting identities artists may express as a result, as will be seen in Chapter 5 on traditional British heavy metal and Chapter 6 on British pop music. It also includes an acoustic analysis of other markers of American English described in Chapter 7 and an examination of how British speakers' perception of /t/ has changed over time in Chapter 8, situating the study within a wider linguistic and cultural context.

5 Heavy Metal Pronunciation

"In heavy metal, there is an intimate connection between the vocals and the instruments, with the voice participating as an equal" (Weinstein, 2000, p. 25). Taking the voice as a structural element of the genre, the opening chapter adopts a sociocultural perspective–following the works of Trudgill (1983) and others–and examines the sung productions of various heavy metal artists, focusing on the realization of word-internal intervocalic /t/. This pronunciation feature, associated with USA-5 model, functions as a sociolinguistic variable indexing either an explicitly American identity or, alternatively, alignment with Pop Song English (PSE) as a standard of popular music. As discussed in Section 2.3.2, the 1980s were a decade of major changes in the music industry and popular music as a whole, and one of the genres at the core of this transformation was what is now known as traditional British heavy metal. With its specific origins as a British genre, crystallizing in the 1970s and reaching global success in the 1980s, heavy metal is relevant to study from a sociophonetic perspective, for its distinctiveness, global popularity, and extreme singing style.

This section provides background on the genre, describing the artists selected to be part of the analysis, before outlining the concrete aims and predictions of the chapter. Then, the methodology of this large scale quantitative corpus study is delineated, before describing the analysis and results. A subsequent discussion on the potential motivations behind the identified style-shift of those heavy metal bands will address notions of audience, genre, authenticity, and overall identity that can be constructed and projected through both linguistic and non-linguistic means.

5.1 Background

5.1.1 Why study traditional British heavy metal?

As one of the most popular genres of the 1980s (Weinstein, 2000), heavy metal provides a useful lens to study style-shifting toward an American model of pronunciation in singing. Heavy metal is under-studied in the diverse literature on the sociophonetics of popular music, although Konert-Panek (2018) examined some pronunciation features of Joe Elliott from Def Leppard, one of the four artists examined in this chapter. In addition to the general lack of research on this genre, the history of heavy metal, from its sociological origins in the 1970s to its popularization in the 1980s, sets it apart as markedly British, and to some extent likely to behave differently than other artists discussed in Trudgill (1983). Heavy metal as a distinct musical genre emerged in the 1970s in the West Midlands region of England (Walser, 2015; Weinstein, 2000). More particularly, the city of Birmingham is central to the origins of the genre, as the birth place of the main three bands credited with creating the genre: Led Zeppelin (as precursors), Black Sabbath and Judas Priest.

In addition to the characteristics of heavy metal singing described in Section 3.2.2, which include high-pitched strong vocals, power, and loudness (Meireles, 2016; Zangger-Borch & Sundberg, 2011), the defining musical features of the genre are loud volumes and distortion, strong electric guitar riffs, and extensive use of power chords (Mesiä & Ribaldini, 2015; Weinstein, 2000). The driving energy of this music, sometimes considered aggressive or angry, is a result of several influences and reflects the particular social and geographical context of its origins. Though Cope (2016) warns against the shortcomings of directly assimilating a genre to a place both he and Harrison (2010) see an important link between the industrial environment of Birmingham and heavy metal music, going as far as to argue that "the sounds of heavy industry" in the city directly inspired the heavy and rhythmic sound of heavy metal. Anecdotally, Tony Iommi, Black Sabbath's lead guitarist, lost the tips of two of his fingers in a workplace accident in the steel factory he worked at when he was young. Because of this, he could not play the guitar as before and adapted by wearing leather caps on his fingers, switching to lighter gauge strings and tuning his guitar a few semitones down to make it easier for him to play, thereby creating a very specific, darker tone that became the signature heavy metal sound.

More generally, the late 1970s and 1980s in the Midlands were a time of drastic social changes: the deindustrialization crisis and recession coinciding with the first period of conservative government under Margaret Thatcher (1979-1983) led to declining wages,

chronic unemployment, and class polarization (Moore, 2016). Early heavy metal, far from a decontextualized type of music, reflected the feelings of rejection, alienation, and anger of a section of the British population, identified as

> [t]he first generation of working-class youth and masculinist rebels who did not have a factory job waiting for them when they left school, and who would therefore constitute a surplus population subject to intensified processes of social control. (Moore, 2016, p. 146)

Heavy metal was therefore born from a combination of particular influences, rooted in a specific historical, geographical and sociological context. It is, in its original form[1], a specific genre that mirrors a distinct identity. On a different level, the beginning of the 1980s marked a turning point for British popular music and heavy metal specifically on a broader national and international scale. The Second British Invasion, and its associated heavy metal variant the New Wave of British Heavy Metal (NWOBHM) marked an increased popularity of British genres in the United States. As such, it is interesting to consider this genre though Trudgill and later researchers' perspective regarding style-shifting toward an American model (whether done actively or by conforming to a standard) versus keeping to their British origins in terms of pronunciation.

5.1.2 Selected artists

The bands under study here were selected because they represent some of the quintessential bands of the genre, but also because they all differ in terms of history, sound and identity, resulting in diverse attitudes toward style-shift and the influence of an American model. Four bands were chosen:

Black Sabbath – Black Sabbath are regarded by most critics and fans as being the founding fathers of the heavy metal genre (Weinstein, 2000; Walser, 1993; Cope, 2016; Harrison, 2010). Formed in Birmingham in 1968, they have helped define British heavy metal and embody some key aspects of the genre's distinctive characteristics, notably a heavier and distorted sound. The original and best-known lineup includes Ozzy Osbourne on vocals, Tony Iommi on guitar, Geezer Butler on bass, and Bill Ward on drums, although they have had various members come and go over the years. According to their website [2] have sold over 75 million records worldwide.

[1] From the mid 1980s onward, the popularization of heavy metal, especially in the United States, led to an expansion of the genre and its fragmentation in multiple subgenres, with their own origins and codes (Weinstein, 2000). Initially, it was an English phenomenon and is considered as such in this thesis.

[2] https://www.blacksabbath.com/home.html. Last accessed: 10/09/2025.

Judas Priest – Formed in 1969, Judas Priest are contemporaries to Black Sabbath. Starting with their second album, *Sad Wings of Destiny* (1976), they firmly established themselves in the heavy metal genre, with some recognizing them as an alternative to Black Sabbath as creators of the genre (Weinstein, 2000). From the suburbs of Birmingham, the most well-known lineup includes Rob Halford on vocals, K. K. Downing and Glenn Tipton on guitar, Ian Hill on bass, and Dave Holland (later replaced by Scott Travis) on drums. Rob Halford's hallmark screams and the strong visual elements of their studded leather biker clothing are emblematic of the genre. They have sold over 50 million records worldwide[3].

Iron Maiden – Slightly later than the previous two bands, Iron Maiden was formed in 1975 in London (though singer Bruce Dickinson grew up in Sheffield) and achieved great success in the 1980s becoming trailblazers in the popularization of the New Wave of British Heavy Metal (NWOBHM). The most well-known lineup includes lead singer Bruce Dickinson, bassist Steve Harris, guitarists Dave Murray, Adrian Smith, and Janick Gers, and drummer Nicko McBrain (recently replaced by Simon Dawson). With over 130 million records sold [4], they are considered the most successful band of the heavy metal genre and have rarely deviated from their original sound.

Def Leppard – Formed in Sheffield in 1977, Def Leppard is composed of Joe Elliott on vocals, Phil Collen and Vivian Campbell on guitar, Rick Savage on bass, and Rick Allen on drums. Though having the least 'metal' sound of the bands, they have become figureheads of alongside Iron Maiden, and have had particular success in the United States. They have sold more than 110 million albums worldwide [5], their fourth album *Hysteria* (1987) selling over 12 million copies in the U.S. alone.

5.1.3 Aims

The origins of heavy metal as a musical genre are firmly rooted in their unique sociocultural and historical context. The 1980s in popular music were a decade of change and saw a rapid popularization of this genre in the U.S., partly though the New Wave of British Heavy Metal (NWOBHM), which may have legitimized the use of British artists' accents in singing. Following Trudgill (1983)'s observations on linguistic modification in song, this chapter investigates the extent to which this notion applies to a

[3] https://judaspriest.com/home/. Last accessed: 10/09/2025.
[4] https://www.ironmaiden.com/. Last accessed: 10/09/2025.
[5] https://defleppard.com/. Last accessed: 10/09/2025.

genre as specific as traditional British heavy metal. The analysis focuses on the realization of word-internal intervocalic /t/ as a sociolinguistic variable which, realized as a flap, may index either a projected American identity or an alignment with Pop Song English (PSE) as a standard in popular music singing. As such, the research questions addressed in this chapter are the following: Do the marked origins and subsequent popularization of the genre outweigh the influence of an American-inspired pronunciation model? How do heavy metal bands navigate and construct their identities in their singing voice pronunciation?

5.2 Methodology

5.2.1 Corpus

The corpus consists of all audio files from all official studio recordings of the four bands under study, for a total of 675 songs recorded from 1970 to 2018. The focus of this thesis is on the 1980s specifically, but works spanning the bands' broader careers were also included for comparison. Table 5.1 gives a breakdown of songs and albums per band, and the full discography can be found in Appendix A.1.1.

Artist	Albums	Songs	Interviews
Black Sabbath	19	174	6
Judas Priest	18	215	8
Iron Maiden	16	153	9
Def Leppard	11	133	8
Total	**64**	**675**	**31**

Table 5.1: *Album, song, and interview counts for the four traditional British heavy metal artists under study.*

Official lyrics for each song were retrieved online through Google search, which directly refers to LyricFind[6], a platform for licensed lyric services including lyric search. They were automatically transcribed (broad phonemic transcription) in IPA with the online software toPhonetics (Mu-Sonic, Ltd., 2013), which, according to its documentation, uses the *Cambridge Dictionary* and the *Carnegie Mellon University Pronouncing Dictionary* as its main sources for transcription, as well as user feedback. In order to have a baseline comparison between the speech and singing voices of these artists, 31 interviews (see the breakdown per band in Table 5.1) were also downloaded from YouTube, over the same time span as the songs in the corpus. They were selected to

[6] https://www.lyricfind.com/. Last accessed: 10/09/2025.

ensure the videos were sufficiently long to provide a good sample of tokens, with the quality and relative lack of background noise and/or music to ensure accurate analysis. When possible, the interviewers were identified and their origins were recorded, but this did not seem to have any impact on final results. Descriptions and links to the videos can be found in Appendix A.1.3.

Following the same methodology, a smaller complementary corpus was also created for two American heavy metal bands, Mötley Crüe and Ratt, for the sake of comparison. Both bands were formed in Los Angeles in 1981 and embody the later American adaptation of British heavy metal, known as glam or hair metal. This was done in order to have a reference point, in addition to Patterson and Connine (2001)'s findings that 98% of intervocalic /t/ in a flapping context are in fact flapped by American speakers. The two bands, formed in 1981 and 1977 respectively, were chosen as archetypal representatives of glam metal, considered to be the direct American-marked successor of traditional British heavy metal. Intending this analysis to be smaller in scale, studio albums released in the 1980s only were included in the corpus, resulting in 5 albums and 51 songs for Mötley Crüe and 5 albums and 50 songs for Ratt. Details of the albums are provided in Appendix A.1.2.

5.3 Analysis

5.3.1 Pre-processing

Recordings

All word-internal /t/ were automatically extracted from the transcriptions using a regular expression. Sequences such as *gotta* or *lotta* were discarded, as they were not considered word-internal, and moreover tend to be more prone to flapped /t/ realizations in British English, even if the phenomenon is marginal (see Section 4.2.1). The resulting list of words containing a /t/ token was manually inspected to identify which tokens could potentially be flapped, according to the rule and characteristics of flapping mentioned in Section 4.1.1, repeated here for reference: "it cannot occur unless it is *preceded* by a vowel or a sonorant (l, n, r), and is *followed* by an unstressed syllable-forming element (vowel, l, r, but not homorganic n)" (Haugen, 1938 as cited in de Jong (2011, p. 2), [original emphasis]). If any doubt remained, the typical American pronunciation was checked in the second edition of the *Longman Pronunciation Dictionary* (Wells, 2000). TextGrids were created for each song and, to make manual segmentation a little easier, intervals corresponding to the number of words containing /t/ tokens were automatically created through a Praat (Boersma & Weenink, 2025)

script, so that all that was left to do was to manually align the intervals to the token in the software. The songs presented a less-than-ideal medium for segmentation however: some recordings date back as far as the 1970s, post-production effects such as reverb and distortion were usually added to the vocals, and above all, vocal and instrumental tracks are mixed down into a single track. This adds a considerable amount of noise to the waveform, spectrogram and the audio in general, making it difficult to isolate specific consonants. Therefore, segmentation of the token also included the two syllables between which the /t/ occurs, that is to say either V_v or v_v contexts. As such, if the token occurred in a two-syllable word like *city*, the whole word was segmented. A double auditory analysis was conducted by myself and another expert in phonetics, to see if the token was in fact realized as a flap or a stop. The very few cases where /t/ was realized as [ʔ] were still annotated as stops, as they were rare and the variant is still typical of a British English accent as opposed to a flap. Tokens on which the two listeners disagreed were re-analyzed, in order to reach perfect agreement. This was done based on the idea that, at least perceptively, the two allophones are distinct enough to be clearly categorized as one variant or the other. Durations for each token were also recorded for further analysis.

Interviews

For interviews, the same process was used to identify word-internal intervocalic /t/ that could potentially be flapped so that their position was known ahead of time. Then, all interviews were carefully listened to and each intervocalic /t/ occurrence was categorized as a flap or a stop.

5.3.2 Statistical analysis

Statistical analysis was implemented in R (R Core Team, 2024), within RStudio (RStudio Team, 2025). All tests were carried out directly from Base R. For modeling, the `glm()` function of the lme4 package (Bates et al., 2015) was used for logistic regression. Plots were created using the *ggplot2* package.

5.4 Results and discussion

5.4.1 Results

Overall results

Overall, among the 1 884 intervocalic /t/ tokens that were identified as corresponding to the flapping rule, 1 156 were marked as flaps, or a grand total of 59.4%. The overall count of tokens and how they were realized is given in Table 5.2. The proportion of flapping for each band, as shown in Figure 5.1, was very variable. Black Sabbath displayed the least amount of intervocalic /t/ realized as flaps (37%), whereas Def Leppard showed consistent flapping 91.4% of the time. Less marked than these two extremes, both Iron Maiden and Judas Priest, has approximately the same distribution of flap /t/ realizations at 55.5% and 50.1% respectively.

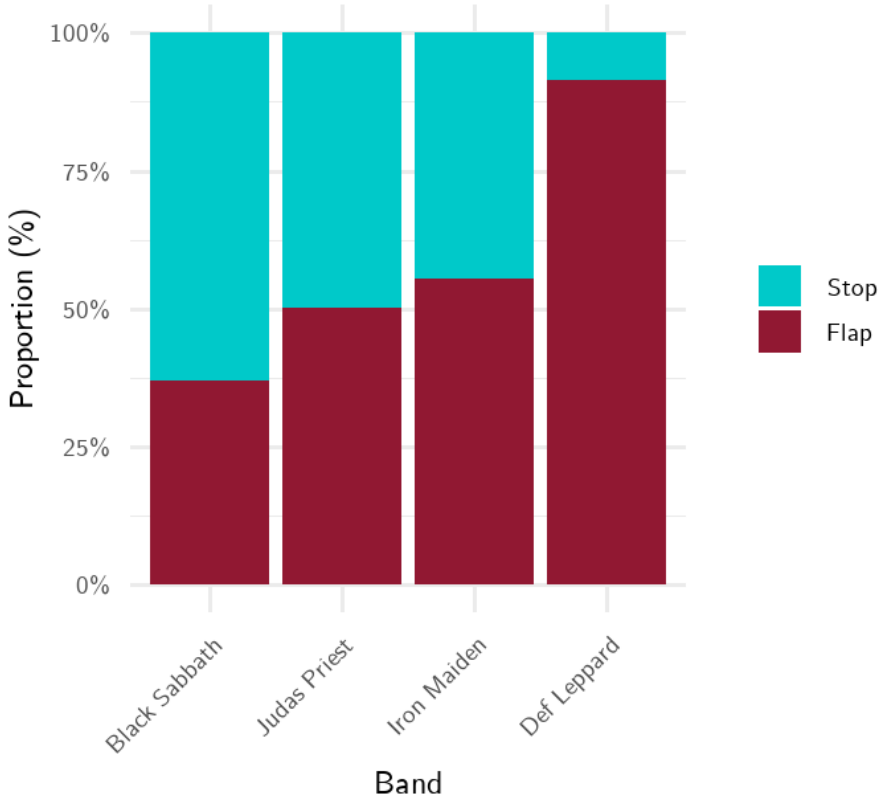


Figure 5.1: *Proportion of intervocalic /t/ realizations in 'flappable' tokens in the sung productions of four bands: Black Sabbath, Judas Priest, Iron Maiden, and Def Leppard.*

A chi-squared test of independence indicated that the likelihood of producing a flap versus a stop depends strongly on the band ($p < .001$). Binomial tests for each

Artist	Flap	Stop	Total
Black Sabbath	114	194	308
Judas Priest	280	278	558
Iron Maiden	261	209	470
Def Leppard	501	47	548
Total	**1 156**	**728**	**1 884**

Table 5.2: *Number of 'flappable' tokens realized as flaps and stops in the sung productions of the four bands.*

band were then conducted to verify if the proportion of flaps within each band differed significantly from chance. This was the case for Black Sabbath and Def Leppard ($p < .001$), as well as Iron Maiden to a slightly lesser degree ($p < .01$). For Judas Priest the flap rate did not differ significantly from chance ($p = 0.966$). No flap /t/ were identified in the auditory analysis of the spoken interviews of the four bands. Regarding the smaller-scale American database of 1980s releases by Mötley Crüe and Ratt, out of a combined total of 278 /t/ tokens, 251 were realized as flaps, or over 90%. More precisely, Mötley Crüe flapped 98% of tokens, and Ratt flapped 84% of them.

Duration

Durations for all /t/ tokens were also included in the data. As a reminder, tokens here refer to both the occurrence of /t/ as well as the syllables preceding and following it (see 5.3.1 for the reason behind this choice). On average, flap /t/ tokens were found to be shorter in duration (592 milliseconds) as opposed to stop /t/ tokens (727 milliseconds). A boxplot illustrating the distribution of token duration according to their realization can be found in Figure 5.2. To keep the data in a format that can easily be interpreted for the plot, no log transformation was applied, but outliers where the duration of the token was greater than 1.5 seconds (1 500 ms) were removed manually ($n = 92$, or about 4.9% of the data). As shown in the plots, flaps tend to occur with shorter durations than stops, consistent with the description of flaps as one of the shortest consonants in English (Iverson & Ahn, 2007; Zue & Laferriere, 1979). The experimental approach in Part III consequently examines the hypothesis that shorter syllable durations may lead to more flapped realizations.

A logistic regression model was fitted to the data, with /t/ realization as dependent variable, and log-transformed duration, band, and their interaction as predictors. The outcome of the model and its syntax are reported in Table 5.3. As expected, the main effect of duration proved to be statistically significant ($p < .001$), with longer tokens less likely to be realized as flaps. The effect of duration on flapping did not differ

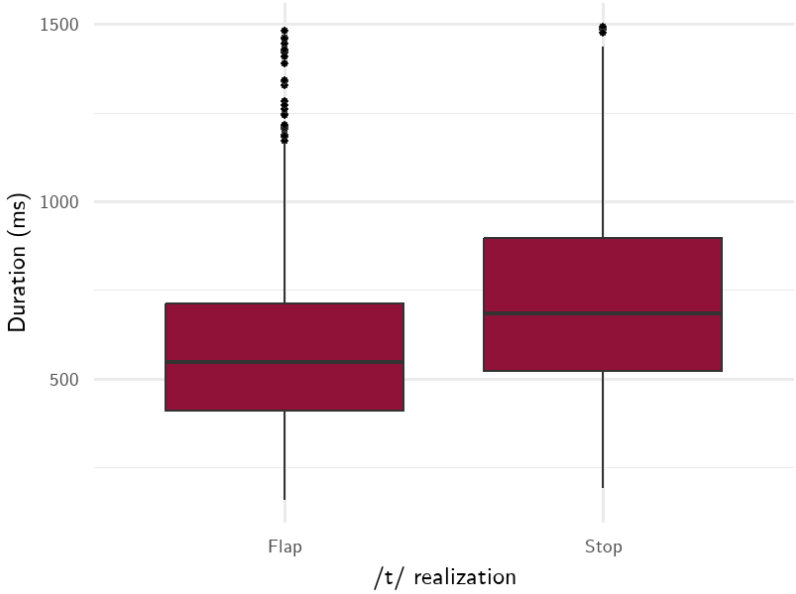


Figure 5.2: *Boxplot of token duration according to /t/ realization.*

significantly by band. Additionally, the model confirms the visual finding in Figure 5.1 that both Judas Priest and Iron Maiden flap more than Black Sabbath ($p < .05$ for both), and more strikingly that Def Leppard flaps considerably more than any other band ($p < .001$).

Diachronic evolution

Although this thesis focuses on the 1980s as a decade of great changes in the music industry, albums from the bands' entire careers (up to 2018) were included in the corpus, in order to get some perspective on a potential evolution of /t/ flapping in singing. Proportions of /t/ realized as flaps for each album under study of each band were therefore included in the analysis and illustrated in Figure 5.3. Each point in time marks a new studio album release. A linear regression line (dotted and dark red) with 95% confidence interval was added to visualize the overall trend in flap proportion across years. Different colored dots indicate the different singers these bands have had through the years where lineup changes have occurred. This is mostly the case for Black Sabbath, which has had five different singers, but there have also been some changes for Judas Priest and Iron Maiden.

5.4.2 Discussion

Despite noteworthy variation in flapping rates between the four British heavy metal bands (as seen in Figure 5.1), all bands display a propensity to realize intervocalic

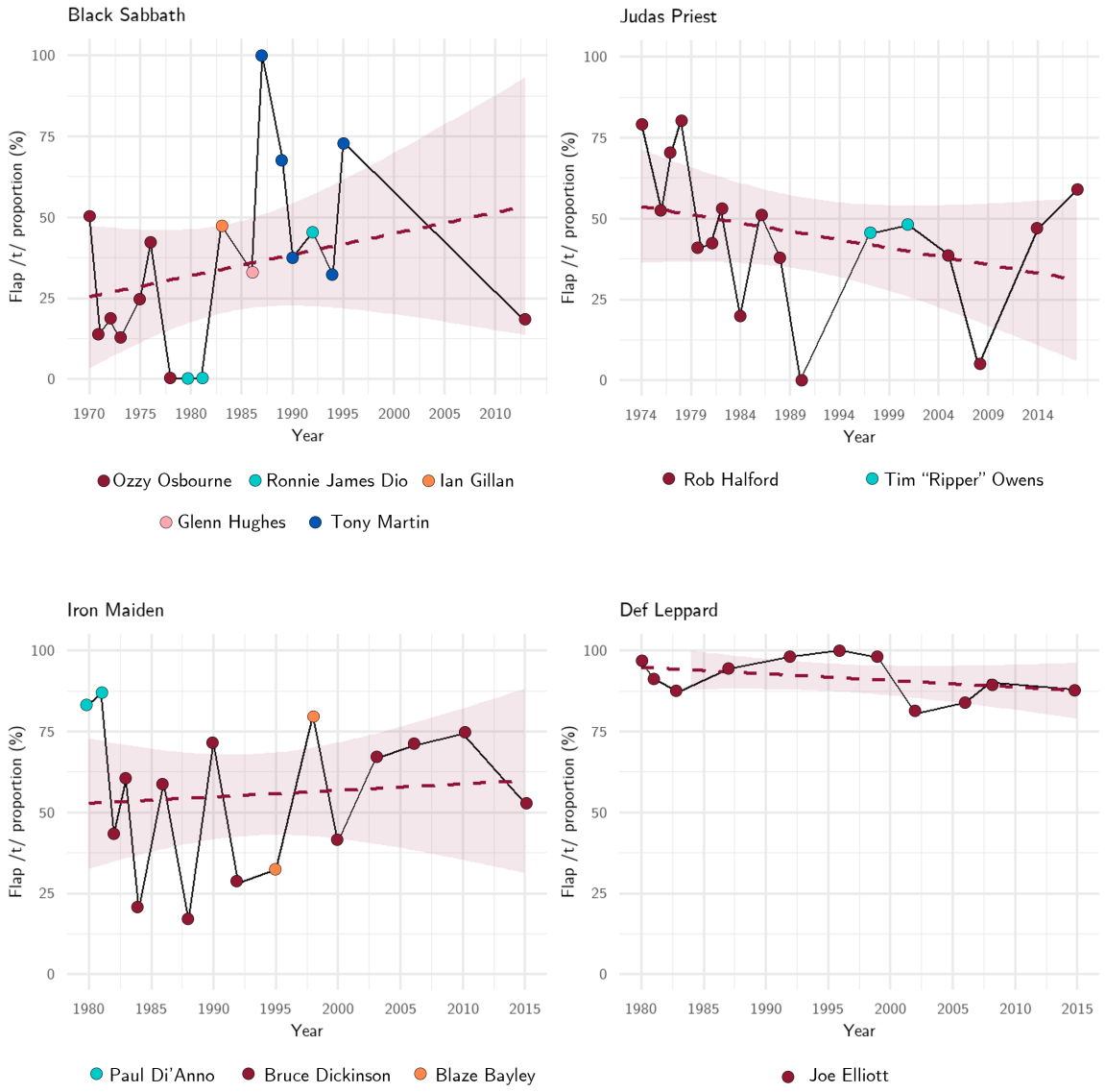


Figure 5.3: *Proportion of flap /t/ realizations per studio album per band, with trend lines.*

Predictor	Estimate	Std. Error	*p* value
(Intercept)	-1.155	0.199	< .001 ***
log(Token Duration)	-1.411	0.337	< .001 ***
Judas Priest	0.585	0.240	< .05 *
Iron Maiden	0.530	0.270	< .05 *
Def Leppard	3.092	0.327	< .001 ***
Duration:Judas Priest	-0.078	0.408	0.847
Duration:Iron Maiden	0.027	0.427	0.950
Duration:Def Leppard	0.603	0.540	0.263

*/t/ Realization ~ log-transformed Token Duration * Band*

Table 5.3: *Output of a logistic regression model predicting flapping as a function of log-transformed token duration and band.*

/t/ as a flap in their singing voice, something which none of them do in the spoken interviews. There is therefore an identifiable modification in their singing versus speaking pronunciation. In other words, there is clear evidence of style-shift in heavy metal bands in spite of their markedly British origins which do not, it seems, entirely outweigh other factors influencing this shift. These findings are in keeping with Simpson (1999)'s claim that "linguistically ... heavy metal has resolutely followed the USA model" (p. 360) The next logical step, then, involves examining the motivations behind this shift, whether automatic or intentional. Using Bell (1984)'s Audience Design concepts, and referring back to the discussion between a more overt "initiative" style-shift toward an American model versus the "responsive" style-shift that reflects a type of pronunciation indexed as standard in popular music (see Chapter 2 and specifically Section 2.3), what can be said more specifically about traditional British heavy metal bands in the 1980s? To answer this question, I will refer throughout this section to four[7] of the five dimensions of language performance identified by Bell and Gibson (2011) and described in Section 1.3.2, namely: identity, the audience, authenticity, and genre.

The American domination of popular music

The main reason put forward by Trudgill (1983, p. 144) to explain British artists' linguistic modification toward an American accent is that "Americans have dominated the field, and cultural domination leads to imitation: it is appropriate to sound like an American when performing a predominantly American activity." Despite the First British Invasion and Second British Invasion, and a certain legitimization of the British accent in popular music throughout the 1980s, this decade nevertheless re-

[7]The fifth dimension, reflexivity, whereby the performance draws attention to itself is more restricted to live performances, which are outside the scope of this research.

mained marked by the domination of the United States in all spheres of popular music, from production to commercialization, the launch of MTV in 1981 providing an example. With regards to heavy metal in particular, the evolution of its status across the decade was particularly striking, and puts into perspective a certain desire for heavy metal bands to appeal to a rapidly growing American audience. Heavy metal became extremely popular in the very beginning of the 1980s with the rise of the NWOBHM, with Iron Maiden and Def Leppard as its figureheads. The heavy metal genre was thus "transformed from the moribund music of a fading subculture into the dominant genre of American music" (Walser, 2015, p. 11). The creation of MTV, an American TV channel specialized in broadcasting music videos, was extremely influential in this popularization, with direct causal links identified between MTV airing heavy metal, sales figures and the acceptance of the genre on mainstream radio (Weinstein, 2000). By the end of the 1980s, the genre accounted for about 40% of all sound recordings in the United States. There is undoubtedly a majority of subsequent American bands (such as Mötley Crüe and Ratt) constituting this figure, but they were nevertheless inspired by the original British bands who already somewhat geared their pronunciation toward this American model.

The four bands under study were very much aware of the commercial appeal of the United States and hoped for success there as the crowning accomplishment of their careers. Def Leppard's debut album, *On Through the Night* (1980), includes the song "Hello America," in which they imagine touring the U.S.:

> Hello America, hello America
> Well I'm takin' me a trip I'm going down to California
> Yeah, I'm gonna try Hollywood and San Pedro Bay
> I'll tell ya what I'm gonna do
> I'm gonna give my love to you
> I'm gonna take you where the lights are bright

It is worth noting that Def Leppard did achieve the American success they were after: their two studio albums *Pyromania* (1983) and *Hysteria* (1987) sold over 10 million copies in the U.S., with five singles from *Hysteria* alone making it to the top 20 (Earl, 2016). However, the fact that they obviously favored an American audience also led to backlash from their British fans, who in fact pelted them with trash and insults during a live rendition of "Hello America" at a show in Reading, England (Phillips & Cogan, 2009). For K.K. Downing, founding member and lead guitar player of Judas Priest, "breaking into the American market with *British Steel* [the band's 6th studio album released in 1980] was key. ... There was this sense that America was where everything was happening" (Downing & Eglinton, 2018, p. 192). He even mentions adapting

their sound to be more 'America-centric' in later album *Turbo* (1986) released with "an obvious focus on the U.S. market" (p. 124). Their single "Turbo Lover" immediately became a hit in the U.S. Similarly, Black Sabbath have also been outspoken about the attraction to the U.S., lead singer Ozzy Osbourne writing in his autobiography they "were so desperate to go to America to get [their] big break" (Osbourne, 2010, p. 89).

Authenticity and genre

The question of authenticity and (sub)genre is also one that is central to the sociolinguistics of performance, and has been mentioned extensively in previous works on the topic (see, among others, Coupland (2011)). Authenticity and genre are both seen as specific characteristics of an artist, for which a certain type of linguistic behavior is expected. For heavy metal, this perspective helps shed light more particularly on the apparent discrepancies that can be seen in the flapping rates of each band. It's important to note that "the heavy metal subculture is noted for being culturally conservative" (Weinstein, 2000) in terms of their approach to the genre conventions, with older bands still today maintaining the popularity and support today. Black Sabbath and Judas Priest were formed slightly earlier than Iron Maiden and Def Leppard, and have both been credited with creating the genre. This automatically grants them a certain status within the genre, and sets them apart as the baseline against which other bands are compared. Taking into account the specific origins of heavy metal covered in Section 5.1.1, it seems logical that these two bands would have lower flapping rates, and therefore keep a certain 'authenticity' in their pronunciation–whether they do so knowingly remains difficult to assert. Regarding Judas Priest, it is widely considered that their first heavy metal album is in reality their second one, *Sad Wings of Destiny* (1976). The first album, *Rocka Rolla* (1974), has been described by band members themselves as "blues-rock," a type of music directly stemming from the distinctly and historically American blue genre (Hatch & Millward, 1987). There is perhaps a relation to Figure 5.3, which shows a clear decrease in flapping rates between these two albums (79.3% and 52.4%, respectively). The subsequent increase in later albums shown in the figure could be due to a strategic turn toward an American audience. With relation to authenticity and genre, Def Leppard's case is probably the most straightforward. The rapid popularization of heavy metal in the 1980s marked a shift in the genre from a traditional subcultural position to an increasingly mainstream one (Earl, 2016), and bands had to adapt. Def Leppard exemplified this by developing a strategy of 'commercial authenticity' whereby they favored a more melodic and pop-like approach to their music. They are in fact usually categorized as 'pop' or 'lite' metal, a subgenre with a typically lighter sound and heavy use of harmony. Drawing from pop music

influences, they have stated that they "were striving to create a musical hybrid" (Halo, 2007), incorporating "all the good elements" from various genres. Going in that direction musically could support the idea of a responsive style-shift for them, where they flap consistently and more than any other band across their whole career (see Figure 5.3), not so much because of intentionally imitating an American model, but because this type of accent was becoming and is now indexed as a standard of pop singing in general. Revealingly, Phil Collen, the band's lead guitar player stated in an interview appropriately titled "Interview with Def Leppard: America's Brits" (Halo, 2007):

> People always used to ask why we sound like Americans [...] Because that is how we learned to be musicians. [...] It is an American accent before you know it. We learned it that way because the stuff happening in England wasn't really homegrown. The pop music was American based blues and R&B.

This suggests that Def Leppard may have adopted Pop Song English (PSE) in song, their use of flapping as a sign of responsive style shift. The consistent high rates of flapping across their whole career also shows precise accuracy, reflecting the rates found for the American bands. In Gibson and Hay (2025, p. 2)'s words, they are "are not 'putting on' an American accent, but simply singing in a way that reflects the majority of the input they have received in similar contexts." Iron Maiden, on the other hand, display a more "conservative" type of authenticity, rarely deviating from their initial musical standard and "embrac[ing] the role identity generated by their early creativity" (Puri, 2014), as the archetypal British heavy metal band they have become. In Bruce Dickinson's words, "Maiden's success was based on reinventing the wheel that drove the heavy metal bus" (Dickinson, 2017, p. 212). In terms of pronunciation then, though they display some influence of the American model, they almost equally stick to their roots, with similar rates of flap and stop /t/ realization.

A note on evolution

From a diachronic perspective, the four bands display different trajectories in their use of flap /t/, as illustrated in Figure 5.3. General trends are difficult to identify, but some general observations can nevertheless be made. While Judas Priest exhibits a gradual decline in flapping rate over time, Black Sabbath fluctuates considerably with a temporary rise in the late 1980s, coinciding with multiple lineup changes. Iron Maiden remains relatively stable around the 50% mark, but there is variation between albums. By contrast, Def Leppard demonstrates remarkable stability, maintaining consistently high levels of flapping across their entire recording history. Changes in

singers are reflected in the variation between albums, especially for Black Sabbath. Interestingly, even though some of the replacement singers were American, notably Ronnie James Dio in Black Sabbath and Tim "Ripper" Owens in Judas Priest, their flapping rates do not seem to form particular peaks in the data, which should be expected from American singers based on the rates found for American bands Mötley Crüe and Ratt. A fact behind some of these disparities is that the number of tokens vary greatly from album to album. For instance, the two Black Sabbath albums from the 1980s featuring American-born Dio as singer (released in 1980 and 1981) only have four and six intervocalic /t/ present in their lyrics overall. They are granted all realized as stops, but the very small number makes it difficult to draw reliable conclusions. Overall, despite occasional identifiable changes in lineup or genre from one album to another, as well as albums with limited tokens, these variable rates also likely correspond to the idea of inconsistency put forward by Trudgill, where "the end product of this language modification is ... by no means entirely successful" (Trudgill, 1983, p. 145).

Conflicting identities

Le Page's third rider in his Acts of Identity model (Le Page & Tabouret-Keller, 1985) refers to the strength of the speaker's motivation in imitating a model group, which may conflict with the speaker's desire to maintain their unique identity. Much like Trudgill (1983) suggested regarding punk-rock music, conflicting motivations toward different models are reflected in British popular music artists whereby the influence of the American model competes with the assertion of a unique British (working-class for punk-rock) identity. This links back to the discussion on the Third Wave of sociolinguistics variation (Eckert, 2012) and how speakers (and in this case, singers) are agents in constructing and negotiating their identities through style. With heavy metal specifically and in the cases evoked above, whether they are actively referencing an American identity or adhering to standards of Pop Song English–or both–the four bands do not necessarily abandon their British identities. Apart from Def Leppard, the inconsistent realization of intervocalic /t/ as flaps hints at the idea of active style-shifting, and therefore an ongoing conflict between two realizations of the same phonetic trait indexing two very different identities, American versus British. Though Def Leppard seems to be more influenced by PSE than an explicit American model, their conflicting motivations may be less geographical in nature and relate more to the clash between adopting the normative accent of their musical style and preserving their sense of personal authenticity. As such, other phonetic traits in all of the bands' productions likely retain their native pronunciation. Being outside the scope of this

study, a quick auditory evaluation of rhoticity in two songs from all albums featuring an English singer of the four bands was carried out, and, at least impressionistically, showed that all these bands seem to retain a non-rhotic pronunciation of postvocalic /r/.

But perhaps more strikingly, some of the ways these bands maintain and put forward their conflicting British identities is not so much in their pronunciation itself, but in other modalities that shape their performances (Bell & Gibson, 2011). Weinstein (2000) insists on the importance of the visual dimension of heavy metal, from album covers to stage effects, and non-phonetic features of British identity can be found in some of the lyrics of the bands as well. Judas Priest, for instance, highlight in various songs their English origins, alluding to the Black Country as they do in "Monsters of Rock" on *Ram It Down* (1988) or in "Deal with the Devil" on *Angel of Retribution* (2005):

> Forged in the Black Country
> Under blood red skies
> We all had our dreams to realize

Their album *British Steel* (1980) was named after the British Steel Corporation, the operating name of the nationalized steel industry in the UK before privatization. In his autobiography, Judas Priest singer Rob Halford mentioned that they were wary of "diluting [their] British heavy metal identity" (Halford, 2020, p. 322) by replacing guitarist and founding member K.K. Downing with a Canadian guitar player. In a thematic analysis of Iron Maiden's lyrics, Spracklen (2017) argues that the band plays a significant role in reproducing and sustaining popular representations of the British Empire. Their work actively constructs an identity closely tied to Britishness and its associated "mythic milieu." For example, live performances of the song "The Trooper" often feature Bruce Dickinson carrying the British flag while dressed in a redcoat uniform evocative of a British soldier of the imperial era, and the band's mascot, Eddie, is often represented with the Union Jack as well. Moreover, war-themed songs such as "The Trooper" and "Aces High" deal with famous British battles and are typically narrated from the perspective of British soldiers. Def Leppard also use the British symbolism of the Union Jack in a lot of their visuals, stage wear and merchandise, and even featured it as the background of their 2005 compilation album *Rock of Ages: The Definitive Collection*. Unlike Iron Maiden's historical approach, however, this could be interpreted as the band playing up their British identity, perceived as unfamiliar or even novel for an American audience. Black Sabbath is the only band which does not display any outward signs of a certain Britishness or British identity, whether it be in their lyrics, appearance, album cover or general band image. That being said, though they realize some /t/ as flaps which they do not do in speech, they also seem to be

the least driven to shifting their pronunciation, and perhaps do not feel the need to insist on their origins or perform their native identity. In fact, Ozzy Osbourne went on to become famous in the United States for his family reality show on MTV, The Osbournes, where his eccentricity as an English man was often pointed out.

5.5 Summary and conclusion

This chapter presented a large-scale quantitative overview of linguistic modification in the singing voice of four traditional British heavy metal bands, through the flap realization of intervocalic /t/ as indexical of an American-influenced pronunciation. A diachronic approach was used to see the potential evolution of this trend beyond the 1980s. Additionally, an extra analysis looking at the duration of /t/ tokens was also conducted, bearing in mind later production experiments (in Part III) where such a factor of flapping will be studied.

Overall results show that this type of modification does occur, to varying degrees, in the sung productions of all the bands under study: they all display the flap variant in their singing, a trait they do not display in their speech. Though diachronic evolution results were somewhat inconclusive, various explanations addressing the audience, genre, authenticity and overall identity of these artists were put forward to explain the observed style-shift. The variability with which each band aligns with an American-like realization of /t/ is examined through the lens of the opposition between responsive and initiative style-shift (Bell, 2002), in a decade of transition in popular music history. Whether the bands under study are aiming at an explicit American accent in their singing pronunciation, or aligning themselves with a broader pronunciation norm that was becoming increasingly standardized at the time, it is interesting to note the prevalence of this tendency, even in a culturally distinct genre such as heavy metal.

6 Pop Music Pronunciation

The Second British Invasion falls within the general context of popular music in the 1980s, a significant decade of transition for British pop music, on an even larger scale than for heavy metal, as described in the previous chapter. Pop acts such as The Police, Duran Duran, or The Human League achieved great commercial success in the US, consolidating the legitimacy of British pop music–and thereby, British pronunciation in pop music–previously acquired in the 1960s with the First British Invasion. Unavoidably, British pop artists thus had to position themselves in relation to the American model, whether by overtly embracing it, subconsciously adapting to it as a pop standard, or actively rejecting it. Within this context, and in a similar way to the chapter on heavy metal, this chapter examines the extent to which pop music artists in the 1980s were subject to style-shift, focusing on the pronunciation of word-internal intervocalic /t/ as a linguistic variable indexing multiple possible attitudes, depending on its realization[1].

First, 1980s pop music is contextualized within the aims of the chapter, before moving on to the description of the methodology, involving the creation of a large database of sung productions and interviews. The analysis carried out on the corpora is described in a third part. Results and a discussion on various artists' attitudes toward an American model follow, using multiple sociolinguistic theoretical approaches, notably Audience Design and Acts of Identity, discussed in Part I.

[1]This chapter presents work that was previously published in Caillol, C.& Ferragne, E. "Does Frankie Go to Hollywood? From American influence to articulatory phonetics: the singing pronunciation of 1980s pop artists", *Anglophonia. French Journal of English Linguistics*. 35. DOI: https://doi.org/10.4000/11qbe.

6.1 Background

6.1.1 Pop music

The difference in meaning between pop music and popular music is addressed at length in Section 1.3.2, but a brief outline of the distinction is as follows: popular music is an all-encompassing umbrella term for many different contemporary commercial genres–including rock, pop, and country, alongside less 'mainstream' genres related to certain countercultures such as punk, heavy metal or hip-hop. Pop music, on the other hand, is a specific genre of popular music characterized by repeated choruses, hooks, short duration, and a basic verse-chorus-bridge structure (Adorno, 1990). While it draws inspiration from various other genres, and is itself composed of many different sub 'types' (subgenres)–such as new wave, art pop, or synth pop, to name a few–pop music is a music genre in and of itself, and will be studied as such in this chapter.

6.2 Aims

Trudgill argues that the First British Invasion legitimized the use of a British accent in pop singing, at least to a certain extent. This section considers whether the Second British Invasion had a similar effect. The first aim is therefore to carry out a large, quantitative analysis of pop music productions in the 1980s to investigate the distribution of /t/ flapping, indexing an American-influenced singing style or Pop Song English.

In examining whether 1980s pop artists conformed to the American model or adopted PSE as their standard singing pronunciation, it is important to recognize the role of sociocultural factors pertaining to the domination of the American music industry and the authority of American music genres. The second aim of this chapter is therefore to explore such factors and use artists' own perspectives to complement these points and gain insight into how they manage and project their artistic identities.

6.3 Methodology

Corpus

Unlike traditional British heavy metal, pop music covers a wide range of different styles, drawing inspiration from a number of musical influences ranging from jazz to punk. Instead of restricting myself to four artists, I wanted the corpus to be as representative a sample of 1980s pop music as possible and large enough to allow robust quantitative

analysis. Though commercial success is not necessarily a defining characteristic of pop music, sales numbers were considered as a starting point for corpus building. For each year in the 1980s (including 1978, 1979 and 1990, as some albums may have been released earlier but topped the charts in 1980 or 1990) the 10 best-selling albums in the UK by male English artists were selected based on a number of criteria. For the sake of comparison with traditional British heavy metal, artists had to be male to avoid any physiological or sex-based variation, as heavy metal tends to be greatly male-dominated. They also had to be English (rather than British) specifically from English parents and not raised abroad[2]. Moreover, albums that were too marked outside the general umbrella of pop (for example, though the band UB40 was very popular at the time and had one album reach the top 10 charts, their distinct genre is reggae, which does not obey the same codes as pop), live albums, and compilations that may have included songs recorded before the decade were all excluded. This was also the case for any albums that overlapped with the heavy metal corpus, such as Def Leppard's *Hysteria* (1987). For the whole period of the 1980s, this resulted in 86 records released by 54 different artists, including names such as Elton John, Wham!, and Depeche Mode.

To identify a representative sample, 350 songs were randomly selected from this database to constitute the final corpus. Each song was individually inspected: if any of the selected songs happened to be an instrumental track, had very few lyrics, lasted less than one minute, or was a cover of an original track by another artist, it was replaced with another randomly selected song. The final sample for analysis thus consists of 350 songs from 85 different albums recorded by 49 artists over a ten-year period (the full list of songs, albums and artists included in the corpus can be found in Appendix A.2.1). A subset of albums and artists is provided in Table 6.1, for reference. It must be noted that in some cases, the same singer can appear twice: once with his band, and once as a solo artist, for instance Phil Collins and Genesis, George Michael and Wham!, or Sting and The Police. They were nevertheless considered as separate artists and included in the database as such.

As described in Section 5.2.1 for the heavy metal data, the lyrics from the 350 songs were also retrieved online through Google Search and LyricFind[3], and automatically transcribed (broad phonemic transcription) in IPA with toPhonetics (Mu-Sonic, Ltd., 2013).

[2] An exception was made for Cliff Richard who was born and lived for the first 8 years of his life in India during the British Raj.

[3] https://www.lyricfind.com/. Last accessed: 11/09/2025.

Artist	Album (year of release)
David Bowie	*Scary Monsters (and Super Creeps)* (1980)
Electric Light Orchestra	*Time* (1981)
Roxy Music	*Avalon* (1982)
Culture Club	*Colour by Numbers* (1983)
Frankie Goes to Hollywood	*Welcome to the Pleasuredome* (1984)
Go West	*Go West* (1985)
Peter Gabriel	*So* (1986)
Pet Shop Boys	*Actually* (1987)
Erasure	*The Innocents* (1988)
Phil Collins	*...But Seriously* (1989)
George Michael	*Listen without Prejudice* (1990)

Table 6.1: *Random subset of artists and albums (one for each year of the decade) included in the pop song corpus.*

Interview data

Chapter 5 looked at four traditional British heavy metal bands with interviews included in the analysis to provide a comparative baseline between speaking and singing pronunciation. This chapter follows the same methodology, with additional parameters included to uncover information about the artists' perspectives and attitudes toward their own pronunciation or their identity as English artists. For each artist in the corpus, three interviews (where available) spanning the whole decade were retrieved from YouTube. They were selected based on the year the interview was conducted–1979 to 1983 for the first, 1984 to 1986 for the second, and 1987 to 1989 for the third–as well as the audio quality of the video. Additionally, three contemporary interviews (from 2010 onward) were retrieved from YouTube for each artist. Links to the videos can be found in Appendix A.2.1.

6.4 Analysis

6.4.1 Pre-processing

Recordings

The same process as described in Chapter 5, Section 5.3.1, was applied to the pop corpus. All word-internal /t/ were automatically extracted from the phonemic transcriptions using a regular expression. Tokens that could not be realized as a flap per the general flapping rule were discarded. The remaining words were kept and /t/ tokens were segmented manually in Praat (Boersma & Weenink, 2025), including the syllable

preceding and following the /t/, as waveforms and spectrograms included instrumental tracks and added vocal effects, which was not suitable for a precise segmentation at the phoneme level. An auditory analysis was then conducted to determine if the /t/ within the token was realized as a flap or as a stop. Durations for each token were also recorded for further analysis.

Interviews

Orthographic transcripts were obtained using YouTube's automatic closed-captioning feature for the six interviews (when six were available) collected for each artist. While the closed-captioning feature is not entirely reliable, it proved enough to achieve two things. First, words including a word-internal /t/ were retrieved using a regular expression, so that they would be easily identifiable for conducting the complementary auditory analysis of artist's spoken pronunciation. Second, all transcripts were imported into AntConc (Anthony, 2022), a corpus analysis software. This allowed to search through the interviews for specific keywords that may indicate some information on the artists' own attitudes toward their work in relation to some kind of American influence (or lack thereof). The following keywords were researched:

- **Keywords related to the U.S.**: United States, America, US, USA, American, MTV, New York, LA, California
- **Keywords related to the UK**: England, English, British, UK, United Kingdom, London, Great Britain
- **Keywords related to pronunciation**: Accent, Voice, Pronunciation, Articulation

The information retrieved from these searches is inherently subjective, and only reflects artists' positions at a given point in time. My intention in including them was observational–due to the celebrity status of these artists, it was difficult to conceive carrying out a systematic sociolinguistic interview with targeted questions. It did nevertheless help put into perspective some of the artists' behavior in their pronunciation and attitudes toward the U.S. that will be covered in the Discussion below.

Statistical analysis

Statistical analyses were implemented in R (R Core Team, 2024), within RStudio (RStudio Team, 2025). All tests were carried out directly from Base R. For modeling, the `glmer()` function of the lme4 package (Bates et al., 2015) was used for logistic regression and `lmer()` when random intercepts were included. Plots were created using the *ggplot2* package.

6.5 Results and discussion

6.5.1 Results

Overall results

Little to no word-internal flapping occurred in the spoken interviews of the 49 different artists included in the final database, though they were not explicitly counted. In sung productions, out of 937 words in which intervocalic /t/ could be flapped, 571 were realized as flaps, or a total of 61% of all tokens. A binomial test confirmed that there was a significant bias in favor of the flapped allophone ($p < .001$) in the overall database. Nonetheless, the degree of flapping varied considerably when examined at the level of individual artists. Flapping rates for each artist are displayed in Figure 6.1. Each artist had an average of 20 tokens analyzed (median = 14), though 3 artists had under 4 tokens. Results for Johnny Hates Jazz, Peter Gabriel and Orchestral Manoeuvres in the Dark should therefore be interpreted with caution as they may not be representative of these artists' real pronunciation trends. Nevertheless, the distribution of flap /t/ realizations shown in Figure 6.1 ranges from 0% of flapped tokens for Heaven 17 (n = 15) to 100% for artists such as Electric Light Orchestra (n = 17) or Nik Kershaw (n = 15). Beyond individual variation, it is interesting to point out that a majority of artists (30 out of 49) flap 50% or more of their tokens, and almost all of them display the pronunciation trait in their singing despite very rare occurrences word-internally in their speech, as noted from the interviews.

Duration

Similarly to the duration analysis in Chapter 5, duration data for all tokens were also examined for the pop corpus. Confirming the previous findings, flap /t/ tokens were on average shorter in duration (542 milliseconds) than stop /t/ tokens (670 milliseconds). A boxplot illustrating the distribution of token duration according to their realization can be found in Figure 6.2. No log transformation was applied, in order to keep the data intelligible for the plot, but outliers where the duration of the token was greater than 1.5 seconds (1 500 ms) were identified and removed manually (n = 28, or about 2.9% of the data). That flapping is consistently shorter in duration than stop realizations of /t/ was expected (Umeda, 1977; Zue & Laferriere, 1979), and quantitative singing data from both heavy metal and pop genres confirm this statement. A binomial generalized linear mixed model was fitted to the data, with /t/ realization as dependent variable, and log-transformed duration as predictor. Unlike the model used in Chapter 5 for heavy metal, I decided to include artists as a random effect instead of a fixed effect

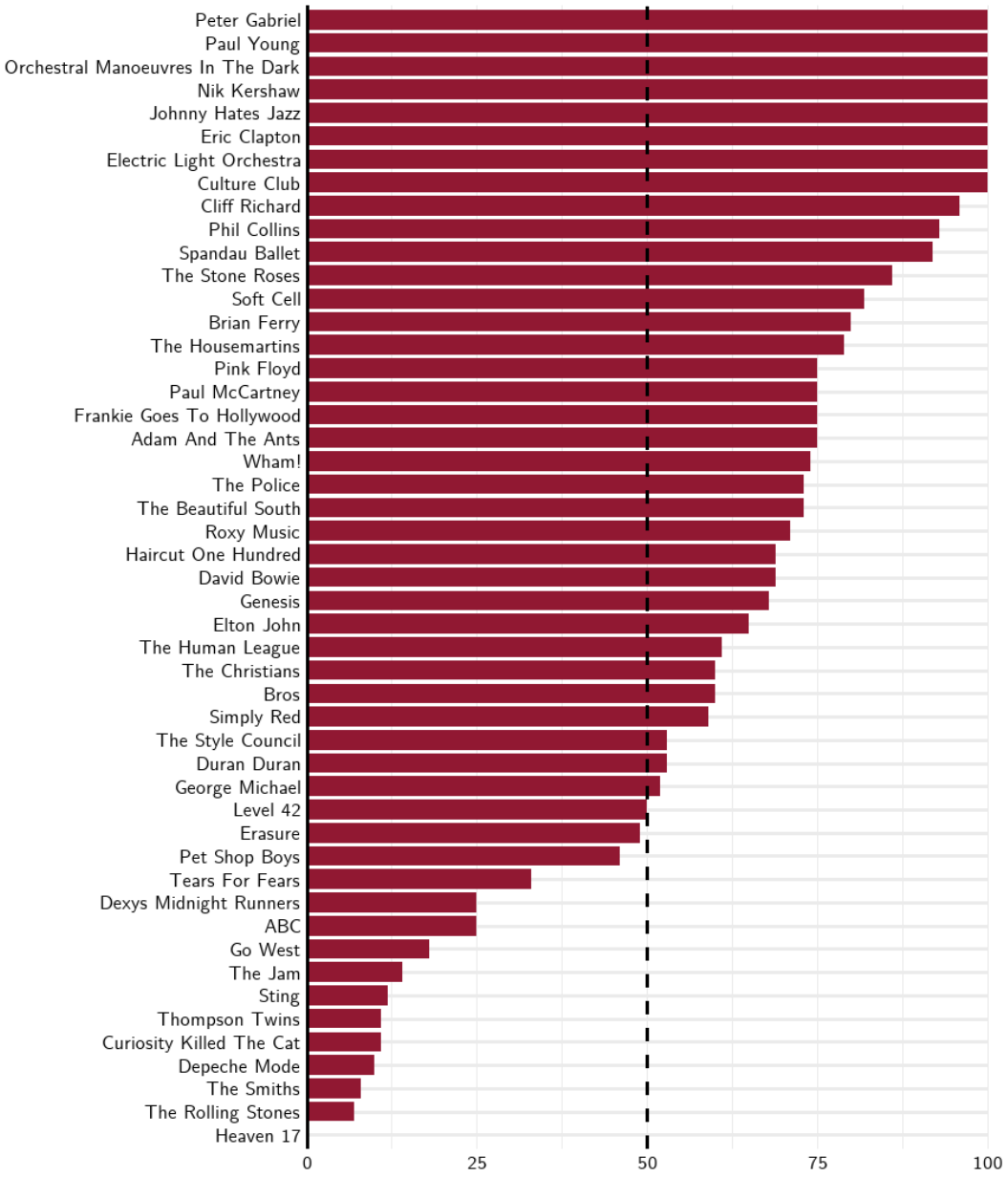


Figure 6.1: *Percentage of flapped intervocalic /t/ for each artist in the pop corpus.*

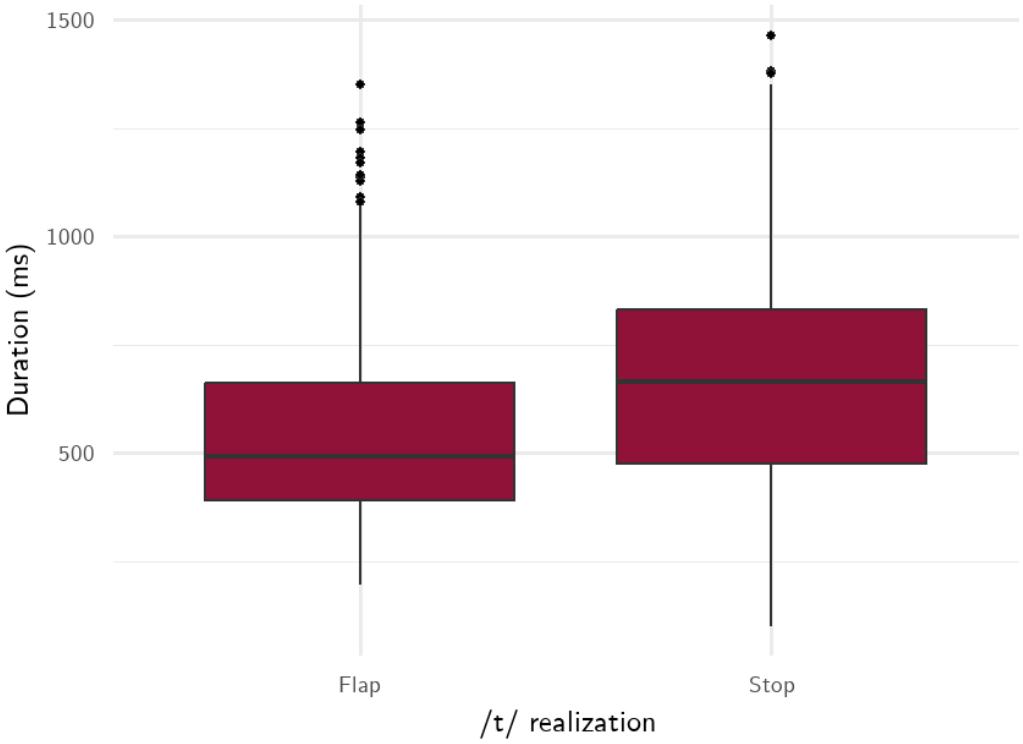


Figure 6.2: *Boxplot of token duration (in milliseconds) according to /t/ realization.*

Predictor	Estimate	Std. Error	*p* value
(Intercept)	-0.214	0.273	0.432
log(Token Duration)	1.437	0.223	$< .001$ ***

/t/ Realization ~ log-transformed Token Duration + (1 | Artist)

Table 6.2: *Output of a generalized linear mixed model predicting flapping as a function of log-transformed token duration.*

to account for the large number of bands and the high amount of variability. Model syntax and results can be found in Table 6.2.

The model confirms that log-transformed duration is a statistically significant predictor of flapping. In other words, flaps are shorter in duration than stops. This is an important factor of distinction between these two variant of /t/, and will be included in the conditions of the experimental approach conducted in Part III of this thesis. To compare to the heavy metal results reported in Chapter 5, Section 5.4.1, a chi-squared test of homogeneity was conducted on stop and flap proportion across genres and determined there was no detectable difference ($p = .966$) in flap and stop duration across heavy metal and pop genres.

6.5.2 Discussion

On the same basis as the discussion of heavy metal results in Section 5.4.2, some of the motivations behind this identified style-shift require further examination. Even if the exact moment of transition from an "initiative" attempt to adopt phonetic features indexed as American, to the enregisterment of these features as the default in pop singing leading to a "responsive" style-shift remains unclear, sociocultural factors undoubtedly play a role. It is necessary to take into account the multiple dimensions of language performance and the influence of a broader sociocultural context on style-shifting.

The American domination of popular music

Trudgill (1983) suggested that the attraction of an American model on the part of British pop singers may be due to what he referred to as 'the American domination' of the music industry. Much like the heavy metal bands discussed in the previous chapter, the United States at the time represented for many artists an ideal of success and accomplishment, where breaking into the American market was a definitive sign of recognition for a popular music artist and brought considerable financial rewards. As Cliff Richard (96% flaps; $n = 26$) put it: "we're going for the next album in America and hopefully if I can get in the top 50 I can feel I'm starting to really crack the market."[4]. Simply put, the U.S. had a far larger music industry at the time. The British Phonographic Industry sets the platinum certification threshold at 300 000 units, whereas platinum is reached in the United States after one million sales, as per the Recording Industry Association of America. These figures reflect vastly different population sizes, and underscore the scale of the U.S. market and the commercial advantages associated with it (Negus, 2011).

In addition to the pure market considerations, a significant motivational force behind style-shifting is likely the cultural authority of American music genres and subgenres, and the inspiration and/or rejection these may elicit in British artist's attitudes. While it's true that the 1980s saw a certain degree of influence of British-branded genres, such as Madchester[5], popular music artists drew major inspiration from American performers (and more specifically *African*-American performers) and music genres ranging from rock and roll to rhythm and blues (Frith, 2001). A closer examination of how productions included in the corpus are categorized into subgenres proves valuable in light of the diversity of 1980s pop productions, as it may help clarify how different artists positioned themselves toward the United States and its cultural influence. For

[4] https://www.youtube.com/watch?v=dvvwqpVvBFk. Last accessed: 11/09/2025.

[5] An indie dance movement that originated in Manchester in the late 1980s. The Stone Roses, included in the database, are one of the major figures of this subgenre.

instance, Eric Clapton has talked extensively about his blues influences and his sense of kinship with African Americans: "I had to just accept that I was using that identification as a way of strengthening my music ... why else would I have responded to the blues and R&B when I was so young and other music didn't reach me that way?"[6] While this statement supports the idea of a responsive shift toward PSE, incidentally, Eric Clapton is also one of the very few artists in the corpus to explicitly mention pronunciation in singing, stating "I don't like a lot of the singing styles that come out of England ... I still like people to sing with an American accent,"[7] demonstrating here a more overt attraction to an American model. In a very similar way to Phil Collen from Def Leppard (see his quotation in Chapter 5), Cliff Richard has also stated that "there would have been no Beatles, no Cliff Richard if it hadn't been for Elvis, Little Richard, Buddy Holly, the Everly Brothers, Chuck Berry. We learned everything from them and even though we didn't quite know what we were doing, we were influenced."[8] Albeit to a lesser extent, other artists acknowledge the influence that America and American genres had on them creatively. This includes Elton John (65% flapping, $n = 112$), "I think America has always been the source of my inspiration, musically anyway,"[9] or Adam Ant from Adam and the Ants (75% flapping, $n = 8$) "America as a country influenced me immensely," and that his way of performing "comes from Springsteen ... he represents everything great about rock and roll."[10] Even with 'only' a 46% flapping rate, ($n = 28$), Pet Shop Boys singer, Neil Tennant, stated that during the 1980s, the duo "felt more part of the New York dance scene than any London scene."[11] There are many more expressions of the same sentiment that can be found in the other interviews (links are provided in Appendix A.2.1), but what seems to show through is that there is, for these popular artists, an attraction toward an American model, whether purely commercial or as a musical inspiration. Adopting American-like features such as a flap realization of intervocalic /t/ can therefore be interpreted as a strategy for appealing to the United States, whether as an initiative style-shift or a responsive one–and in most cases–a mix of both.

Overshooting

Bell and Gibson (2011) mention overshoot[12], among other phonetic processes, as an indication of performed stylization of language use: it is more likely to appear when

[6] https://www.youtube.com/watch?v=3pLGBC1rzwo. Last accessed: 11/09/2025.
[7] https://www.youtube.com/watch?v=sizOT8JRRrY. Last accessed: 11/09/2025.
[8] https://www.youtube.com/watch?v=ri8P4Qvy9NI. Last accessed: 11/09/2025.
[9] https://www.youtube.com/watch?v=KDiPqIiJivQ. Last accessed: 11/09/2025.
[10] https://www.youtube.com/watch?v=05hVdIz9U9s. Last accessed: 11/09/2025.
[11] https://www.youtube.com/watch?v=0zryqgvkrqo. Last accessed: 11/09/2025.
[12] Not to be confused with articulatory overshoot, described in Section 4.1.1.

artists are targeting a model variety they do not use in their daily life (in this case, when speaking). Trudgill (1983) also refers to this as a form of hypercorrection, whereby artists wrongly over-apply a rule to contexts where it does not occur in the target variety. In his work, Trudgill (1983) notably mentions rhoticity as a recurring case of overshoot/hypercorrection. Singers, wanting to imitate an American accent, and able to identify that it is rhotic, over-apply rhoticity and introduce linking /r/ in some of their lyrics: citing for example the Kinks singing "Ma and Pa" as [maːr ən paːr] in "Sunny Afternoon."[13] Though it is difficult to systematically identify cases of overshoot in the corpus, at least two stood out during the analysis, pointing to the conscious adoption of an American accent as opposed to using an enregistered norm. The first example can be heard in the song "I'll be your Shelter" from the album *London 0 Hull 4* (1986) by Cheshire-based band The Housemartins. In it, singer Paul Heaton consistently flaps the /t/ in the word *shelter*, over three repetitions in the song. Although the phonological context would allow for flapping as per the general flapping rule, this word is in fact an exception: it would rarely, if ever, be flapped by a native AmE speaker. There is no potential voicing indicated in the *Longman Pronunciation Dictionary* (Wells, 2000), and a brief auditory examination of 30 tokens of *shelter* said by various native AmE speakers using YouGlish [14]. Another striking example is the way ABC's lead singer, Martin Fry, originally from Manchester, pronounces his very own name in a spoken part of the song "The Look of Love, Pt.1" from *The Lexicon of Love* (1982). In interviews, he introduces himself as ['maːtɪn], but in this song, he can be heard saying ['maːrɪn][15] when referring to himself, effectively modifying the pronunciation of his own name toward what might seem like an American accent. While noteworthy that he flaps his own name, it is even more striking that it is very unlikely an American speaker would flap the intervocalic /t/ in the name Martin, most often pronounced as a glottal stop (though the *Longman Pronunciation Dictionary* suggests /'maːrtən/ in AmE). Gibson (2023) hypothesizes that in the early 21st century, as American-influenced pronunciation became increasingly normalized into PSE, this tendency toward hypercorrection/overshoot decreased. These few examples might suggest that in the 1980s, the process of standardization of PSE is underway, but not yet fully realized.

[13] Trudgill (1983) notes that 'Ma' is a simple case of intrusive /r/, compared to the "hyper-american" (p. 149) /r/ in 'Pa.'

[14] YouGlish https://youglish.com/ (last accessed: 11/09/2025), is a website that lets users hear how words are pronounced in real-world contexts by accessing them through a large database of YouTube videos.

[15] Compare the first few seconds of https://www.youtube.com/watch?v=jpfg0cJKRD8&t=2s to https://www.youtube.com/watch?v=cNEdxZURTaI at 2:42. Videos last accessed: 11/09/2025.

Going local?

In contrast with the examples cited above, some of the artists with the lowest rates of flapping might very well be attempting to "go local" and embrace some sort of personal authenticity by actively maintaining or even emphasizing their native accents. The Jam (14% flapping rate, $n = 14$), for instance, exemplify a band whose musical identity was deeply rooted in Britishness, both through the mod revival[16] aesthetic they put forward and their class-conscious lyrics. This likely influenced their choice to avoid an overt American-like pronunciation of intervocalic /t/. This is also the case for The Smiths (8% flapping rate, $n = 12$). When guitarist Johnny Marr was asked about the name of the band, he stated the reason they chose a generic last name was because they "felt it was about time that a pop group identified themselves with normal people, and [they] wanted to be unconventional in rock and roll terms"[17]. This statement can be interpreted as a direct rejection of the (American) norms linked to pop and rock genres, and a wish to engage with their British roots in "an England that's fighting very hard to exist"[18] While the attitudes of these bands were more geared toward embracing their origins, some artists deliberately situated themselves in opposition to an American model. The only band to show a 0% flapping rate ($n = 15$), Heaven 17, is explicit about not being interested in playing live or touring the U.S. in one of their interviews: "we didn't want to turn into a stadium rock band ... we certainly turned down a really large amount of money to not play America"[19] This association of the U.S. with 'selling out' or at least losing part of an artist's identity and authenticity is also echoed in an interview of Boy George from Culture Club:

> A lot of English bands, when they started to feel success in America they went and they cut off their hair, they took off their false eyebrows and basically made a real compromise but with Culture Club my attitude was go there and give them some culture, you know, bring them some real English eccentricity[20].

However, in the corpus, Culture Club actually flap all their /t/ tokens. This may be due to the fact that there were only five tokens in their sample, but perhaps their idea of showing their English identity to an American audience did not necessarily involve their pronunciation–likely to be influenced by in that case, but clearly demonstrating acts of "conflicting identities" as put by Trudgill (1983).

[16] Mod revival is a subculture that started emulating the aesthetic of the sixties Mod culture in the early 1980s.
[17] https://www.youtube.com/watch?v=2ACFfRmXWsE. Last accessed: 11/09/2025.
[18] same as above.
[19] https://www.youtube.com/watch?v=rxGUuGxGV3Y. Last accessed: 11/09/2025.
[20] https://www.youtube.com/watch?v=fxVjt1e-QPY. Last accessed: 11/09/2025.

Heavy metal versus pop

As in the previous chapter on heavy metal /t/ pronunciation, the average flapping rate found (59.4%) was very similar to the one reported for the pop music corpus (61%). This suggests that despite different origins and sounds, the driving force toward an American-like model–whether overt or as an enregistered norm–is similar for these genres. The same sociocultural pressures likely act on artists in both genres, as evidenced by comparable attitudes of attraction and rejection of the American model in relation to performance, subgenre and authenticity, audience, etc. Yet, this apparent similarity in style-shift may hide an alternative explanation: the high flapping rate in heavy metal may be more closely governed by articulatory constraints than in pop, given heavy metal's wider pitch range, greater volume, higher vocal intensity, and overall increased 'power,' as will be addressed in Part III.

6.6 Summary and conclusion

This chapter investigated the pronunciation of British pop artists in the 1980s, with a particular focus on the realization of intervocalic /t/. Drawing on a database of 350 songs from 49 artists complemented by interviews, the analysis revealed that flapping was highly frequent in sung production (61%), despite being extremely rare in the same artists' speech. One limitation of this study is that the artists' speech was not analyzed systematically, but was instead inferred from auditory impressions of interviews. Nevertheless, although some artists occasionally produced flaps in speech, the occurrence was rare enough to make its presence in singing noteworthy. The prevalence of flapping in these artists' sung productions confirm that it functions as a marker of style-shift in British pop singing. At the sociocultural level, the results suggest that British pop artists of the 1980s negotiated their identities in relation to an American model, whether by adopting American-like pronunciation as a means of accessing the U.S. market, by explicitly aligning with American musical genres and traditions, or by resisting such associations in the name of authenticity and 'local' identity. Interviews provided evidence of both attraction to and rejection of the American model, with most artists acknowledging the authority of American genres even when articulating a desire to maintain British distinctiveness.

In relation to the chapter's aims outlined in Section 6.2, the 1980s appear to represent a period of transition–both in the broader domains of music production and commercialization, and in pronunciation style. This shift involved a movement away from the deliberate adoption of an American accent as a marker of success, popularity, and affiliation with particular subgenres, toward the more unconscious, reflective use

of PSE as a supra-local singing norm.

7 A Complementary Acoustic Analysis

Whether due to direct American influence, the result of a normalized standard across anglophone popular music in general, the outcome of specific genre inspirations, or the desire to project a certain identity, previous chapters have shown that style-shift in British singers is a widespread phenomenon. Nevertheless, while a single pronunciation feature such as intervocalic /t/ is distinct enough for such a general conclusion to be drawn, other features may help consolidate this idea. This was pointed out by Simpson (1999)'s USA-5 model, which provides five distinctive American pronunciation features. Choosing to focus on intervocalic /t/ was a methodological choice, in order to analyze this feature exhaustively in all its manifestations, but also because of the assumption that it would be difficult to carry out detailed acoustic analyses on pre-existing mixed track recordings.

Nevertheless, taking into account the importance of the USA-5 model, I decided to carry out a comparative acoustic analysis across different genres and artists, if only to verify some of the conclusions drawn from the analyses carried out in previous chapters. This chapter[1] is therefore a complementary, exploratory look into vowel analysis on pre-recorded tracks, assessing how we may adapt traditional acoustic analysis to a suboptimal context, using various tools and testing their reliability. This chapter starts by giving some background into acoustic studies of the singing voice, before presenting the small sample of artists, songs and pronunciation traits selected for analysis. It then goes into the methodology and analysis before moving on to results and their

[1] Caillol, C., & Ferragne, E. (2022). Américanisation de la prononciation de l'anglais en voix chantée : le cas des voyelles de TRAP-BATH et LOT. In XXXIVe Journées d'Études sur la Parole (JEP).

discussion.

7.1 Background

7.1.1 Acoustic analysis of the singing voice

Studying the acoustics of the singing voice in recorded popular music songs poses two major problems. The first has to do with the songs themselves, and how they were recorded. While studying acoustics is done without much difficulty in operatic or choral singing (Marshall et al., 2024; Sundberg, 1995), since vocals-only performances and recordings are available, this is less the case for popular music. Most, if not all, of the pre-recorded songs studied in the previous chapters took the form of a final mixed track where instruments and vocals have been combined. Granted, some of the works mentioned in Chapter 2 examined style-shift in popular music using acoustic analysis methods, but they did so within a limited scope. Most of the studies cited in Section 2.1.2 relied on impressionistic results from auditory analysis, sometimes paired with visual observation of spectrograms as in Gibson (2023). In other cases, researchers worked with isolated vocal tracks available online (Caillol & Ferragne, 2019; Konert-Panek, 2017b), or even recorded artists directly, as Sundberg and Thalén (2015) did in their comparison of neutral and belt singing styles.

The second issue linked to analyzing the acoustics of the singing voice in popular music recording is whether or not formant extraction can be done, and whether they are robust enough to gather accurate data. This is questionable first and foremost as speech and singing differ, so direct comparison is not methodologically possible, but also because most of the vocals in the recordings under study have been altered in post-production with effects such as reverb, delay, or phaser. Furthermore, the songs under study date back to the 1980s and present issues in terms of recording quality, with technology that may be considered out of date today. Though formants may usually be considered robust enough to carry out such analyses, this remains to be tested empirically.

7.1.2 Selected artists

Six artists were selected for this analysis. There were four British artists: Duran Duran and Phil Collins represented the pop genre, and Iron Maiden and Judas Priest for traditional heavy metal. As a baseline for comparison of their pronunciation, two American pop artists were also included in the study: Billy Joel and Bruce Springsteen. Iron Maiden and Judas Priest have previously been presented in Section 5.1.2. The others

were chosen as quintessential representatives of their genre, based on their popularity in the 1980s and how prolific they were in their music production during the decade, so that the overall number of songs under study would be adequate.

Duran Duran – Duran Duran is an English pop rock band (associated with New Wave and New Romantic music genres) formed in Birmingham in 1978. Its lineup throughout the 1980s was composed of keyboardist Nick Rhodes, bassist John Taylor, guitarist Andy Taylor, drummer Roger Taylor and singer Simon Le Bon. Le Bon was born in Hertfordshire but later attended the University of Birmingham. According to their official website[2], they have sold over 100 million records over the course of their career, including 20 UK Top 20 singles and 18 U.S. Hit singles.

Phil Collins – Born in south-west London, Phil Collins started his career as drummer and later lead singer of the progressive rock band Genesis. He also had very successful solo career as a pop singer in the 1980s, which is what will be studied here. His solo discography includes 8 albums that have sold an estimated 150 million copies worldwide[3].

Billy Joel – is an American pop rock singer-songwriter. He has released 12 studio albums, sold over 160 million records worldwide, and is on par with Michael Jackson as the sixth best-selling solo artist in the United States, according to the Recording Industry Association of America[4]. He rose to fame in the 1970s and consolidated his popularity all throughout the 1980s.

Bruce Springsteen – Perhaps one of the most iconic American singer-songwriter in his image and the themes he tackles in his songs of working-class America (Frith, 1988), Bruce Springsteen has sold over 140 million albums worldwide including 70 million in the United States alone[5]. He rose to fame in the 1980s, especially with the release of his 1984 album, *Born in the USA*.

7.1.3 Selected vowels

Referring back to the USA-5 model of salient pronunciation shifts in British popular music artists, as described by Simpson (1999) and illustrated in Table 2.1, two vocalic features were selected for analysis: the TRAP-BATH split and LOT unrounding. The

[2] https://duranduran.com/. Last accessed: 10/09/2025.
[3] https://philcollins.com/. Last accessed 10/09/2025.
[4] https://www.riaa.com/. Last accessed: 10/09/2025.
[5] https://brucespringsteen.net/. Last accessed: 10/09/2025.

TRAP-BATH split refers to a sound change in which, in certain contexts, the short front vowel /æ/ (as in *cat* or *back*) changed to a longer and back /ɑː/, the vowel found in the BATH lexical set, in words like *ask* or *dance* (Wells, 1982a). This distinction developed primarily in southern varieties of British English (BrE) but did not occur in American English (AmE), where BATH words continue to be pronounced with /æ/. As a result, dance is pronounced [dɑːns] in BrE but [dæns] in AmE. LOT unrounding is another pronunciation feature that distinguishes BrE from AmE. For words in the LOT lexical set (*stop*, *dog*, *hot*, etc.), BrE typically uses a rounded vowel [ɒ], while AmE tends to pronounce them with the unrounded vowel [ɑ]. Wells (1982a) notably refers to this trait as an "indigenous American development" (p. 245), [ɑ] realizations of LOT being extremely rare in present-day BrE.

7.1.4 Aims and predictions

Given the potential issues associated with acoustic analysis of popular music, this chapter has two main aims. Firstly, a comparative analysis between genres (heavy metal and pop) and nationalities (British and American) will be carried out, in order to have a glimpse of style-shifting processes in sung vowel sounds, in addition to the consonant /t/ that has been examined so far. The second aim is to explore potential solutions to the issues described in Section 7.1.1 above, using a music source separation tool (described in Section 7.2.1) and verifying the accuracy of conventional methods of formant estimation and extraction. As such, the following hypotheses are proposed:

Hypothesis 1 Based on the similar general pronunciation trends identified in the heavy metal and pop corpus analyses, there should be a general trend toward the American-like pronunciation of certain vowels for both genres: a shortening of BATH vowels toward a TRAP realization, and an unrounding of the LOT vowel.

Hypothesis 2 The use of a vocal extraction tool will significantly improve the feasibility of acoustic analysis, yielding sufficiently reliable formant extraction results.

7.2 Methodology

7.2.1 Corpus

Song selection

All songs (282) from all studio albums (excluding compilations) released between 1980 and 1989 (29) by the six artists under study were included in the corpus. Table 7.1 shows the album and song count per artist, and album names can be found in Appendix A.3.1. Similarly to the heavy metal and the pop corpora, corresponding lyrics from

Artist	Albums	Songs
Billy Joel	5	48
Bruce Springsteen	4	54
Duran Duran	5	43
Phil Collins	4	40
Judas Priest	6	54
Iron Maiden	5	43

Table 7.1: *Album and song counts for artists in the acoustic analysis corpus.*

each song were retrieved online through LyricFind[6] via Google search. They were automatically transcribed in IPA through the online software toPhonetics (Mu-Sonic, Ltd., 2013). The American artists' lyrics were transcribed both in American English and British English to make vowel identification easier: a regular expression using the IPA symbol corresponding to the vowels under study was used and looped through all songs to extract corresponding words. Five songs per artist (30 total) were hand-checked to ensure the transcriptions were error-free.

Vocal track isolation

As mentioned above, conducting reliable studies on mixed tracks is difficult, which explains the small number of large quantitative analyses on recorded music. Spleeter (Hennequin et al., 2020) is an open source music source separation tool developed by Deezer Research[7], the research division of the French commercial music streaming service Deezer. Spleeter uses deep learning models trained on large datasets of musical recordings to decompose mixed-track audio into distinct sources, most commonly separating the lead vocal track from the instrumental accompaniment. Among

[6] https://www.lyricfind.com/. Last accessed: 10/09/2025.
[7] https://research.deezer.com/. Last accessed: 10/09/2025

similar tools, a study by Ansal and Ancy (2021) found that it had the best results at the time of publication[8] for source separation, including signal-to-distortion ratio (SDR) and signal-to-artifacts ratio (SAR). Spleeter was executed in the Google Colab environment using scripts provided in its official GitHub repository[9] (last accessed 14/09/2025).

In this study, the tool was used in an exploratory capacity to evaluate its effectiveness for isolating vocal tracks and to determine whether such separation could benefit acoustic analysis of mixed-tracks recorded music. While Spleeter successfully extracted vocals, the process inevitably reduced audio quality and introduced some artifacts. Its performance was somewhat inconsistent across songs and artists, likely depending on the presence of effects in the original mix, which can interfere with the model's ability to achieve a clean separation. Nevertheless, visual comparison of several spectrograms in Praat (Boersma & Weenink, 2025) before and after the separation (see an example in Figure 7.1) confirmed the potential benefits that could be drawn from using such a tool for further acoustic analysis, including formant extraction.

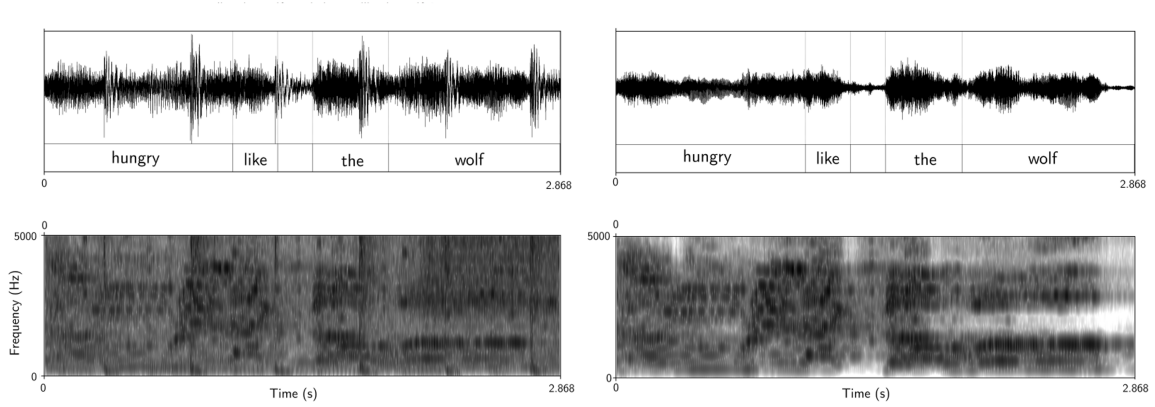


Figure 7.1: *Comparison of a short excerpt from the song "Hungry Like the Wolf" by Duran Duran. On the left, the waveform and spectrogram 'as is' from the mixed instrumental and vocal track and on the right, the resulting waveform and spectrogram of the vocal track only after source separation using Spleeter. Settings were kept at their default values in Praat.*

[8] This acoustic study was carried out in 2022, at the beginning of PhD. With the rapid development of AI, there is no doubt that a better version or better tool has been created since then, but I will nevertheless report the results found using Spleeter.

[9] https://github.com/deezer/spleeter

7.3 Analysis

7.3.1 Pre-processing

To isolate occurrences of the TRAP-BATH split, all words belonging to these two lexical sets were automatically extracted from the transcribed lyrics files, using a regular expression to identify all occurrences of the vowels /ɑː/ and /æ/. A random sample of 100 occurrences of each vowel was retrieved for all six artists, in the transcriptions corresponding to their respective varieties of English. Judas Priest and Phil Collins had less than 100 occurrences of /ɑː/, therefore all tokens of this vowel were kept for these two artists, 81 for Judas Priest and 66 for Phil Collins. The same number of /æ/ occurrences were therefore selected in order to maintain a balanced comparison. A total of 547 tokens were therefore obtained for both vowels for all participants.

The same process was carried out for LOT unrounding analysis, identifying all words containing /ɒ/ in the transcribed lyrics of the six artists. For this specific extraction, the British English automatic transcription of American artists Billy Joel and Bruce Springsteen was used, since the LOT vowel is usually transcribed /ɑː/ in American English. A random sample of 150 occurrences of the vowel was retrieved for each artist, resulting in 900 tokens to analyze.

7.3.2 Acoustic analysis

The resulting 1 954 vowel tokens were manually segmented in Praat (Boersma & Weenink, 2025). F1 and F2 were semi-automatically measured over the whole duration of the vowel using the `cp_formants` Praat program (Ferragne, 2025). This program overlays Praat's formant estimation over the spectrogram, and, if needed, lets the user adjust it manually until it aligns coherently with the spectrographic representation Figure 7.2 shows a screenshot from the GUI, with the estimated formant tracks in red. Mean values in Hertz of F1 and F2 were then extracted for each vowel. A total of 31 tokens were excluded from the analysis because the formants were not visible in the spectrogram.

7.3.3 Reliability analysis

Despite the associated difficulties mentioned in Section 7.1.1, formant estimation and extraction in sung passages is not a new technique. In the past, it relied mostly on visual inspection of the spectrogram (Sundberg, 1970), and more recent experimental setups have used impedance measurement techniques at the recording stage, using a

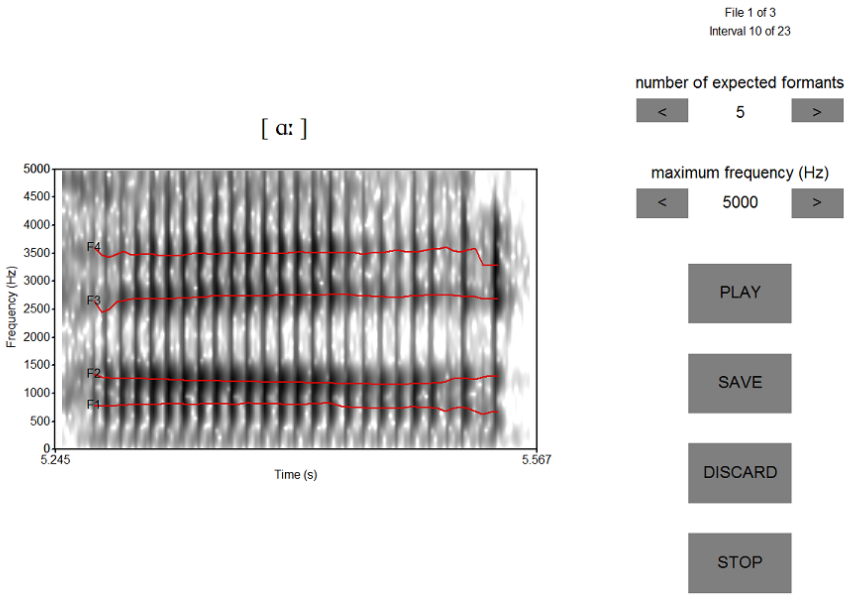


Figure 7.2: *Screenshot from the cp_formants GUI, showing the imposition of the estimated formant track in red over the corresponding spectrogram. Image retrieved from https://github.com/emmanuelferragne/CminR-Praatik.*

broadband acoustic current source to excite the vocal tract just outside the lips. The resulting magnitude response is then used to calculate the resonance frequencies and thus avoid issues associated with post-hoc formant analysis (Epps et al., 1997; Henrich et al., 2011). Other acoustic studies of singing simply use formant measurements without much consideration for whether they may accurately represent underlying vocal tract resonances. The songs examined here were not recorded in a controlled experimental context and since further processing was used to isolate the vocals from the rest of the instrumental track with Spleeter, verifying the reliability of formant extraction in this study was necessary.

MFCC (Mel-Frequency Cepstral Coefficients) coefficients were therefore computed in MATLAB for all vowels in the corpus. The signal was resampled at 16 kHz, and a 30 ms Hamming window with a 20 ms analysis step was applied. Thirteen coefficients were retained and averaged for each vowel. Since it was carried out as a secondary step to the formant analysis, more details on the methodology are provided in the dedicated Results section.

7.3.4 Statistical analysis

Statistical analysis was implemented in R (R Core Team, 2024), within RStudio (RStudio Team, 2025), unless otherwise stated. All tests were carried out directly from Base R and the *car* package (Fox & Weisberg, 2019), and plots were created using the

`plotVowels` command from the *phonR* package (Johnson, 2015).

7.4 Results and discussion

Results for the two features under study are presented separately, but Figure 7.3 includes all three vowels under study for each artist. Ellipses encompass ±1 standard deviation around the group means.

7.4.1 TRAP-BATH

A Mann–Whitney test was conducted to compare the F2 values of /æ/ and /ɑː/. This non-parametric test was chosen over a t-test because some BATH tokens were expected to be realized with a TRAP-like quality, potentially skewing the distribution and violating the assumption of normality. Because the hypothesized contrast between the two vowels of the split lies along the front–back dimension, the analysis is limited to F2. No significant difference was found for most artists: Billy Joel ($W = 4090$, $p = .093$), Bruce Springsteen ($W = 3739$, $p = .413$), Iron Maiden ($W = 3823$, $p = .115$), and Duran Duran ($W = 3578$, $p = .820$). The difference is marginally significant for Judas Priest ($W = 2159$, $p < .05$), but only Phil Collins shows a clearly significant difference in F2 between the TRAP and BATH vowels ($W = 2346$, $p < .001$).

A Levene's test was conducted for Phil Collins to assess the homogeneity of variance between the two vowels. The significant result ($F = 8.57$, $p < .01$) indicates that the variance in F2 for BATH-type words is not equal to that of TRAP words, which may conceal multiple realizations of the BATH vowel in this artist's productions. An auditory analysis was therefore carried out on the 10 occurrences of /ɑː/ with the highest F2 values (between 2 007 and 2 148 Hz) and the 10 tokens with the lowest F2 values (between 755 and 1 297 Hz) to test this hypothesis. For the occurrences with the highest F2 values, all vowels expected to correspond to BATH were in fact realized with the open front [æ] vowel of TRAP. For the occurrences with the lowest F2 values, 7 out of 10 vowels aligned with a [ɑː] realization. An auditory examination of the 10 occurrences of the BATH vowel with the lowest F2 values (between 1 125 and 1 523 Hz) was also carried out for Judas Priest, whose statistical test suggested a marginally significant difference. Although the realization of two vowels in this sample tended toward [ɑː], the low F2 values of the other occurrences appear instead to be associated with distortion effects applied to the voice during recording.

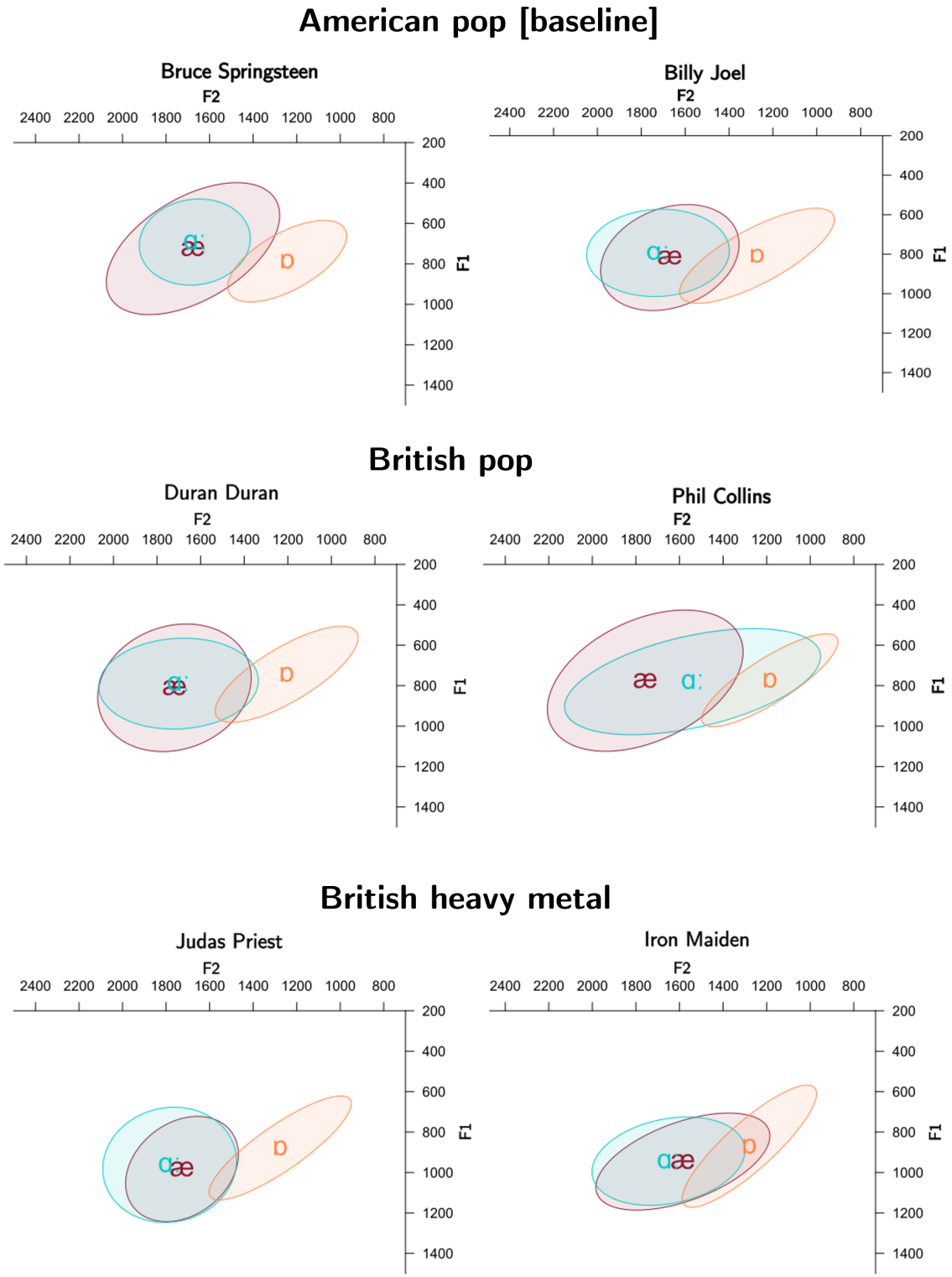


Figure 7.3: *Average F1 and F2 values for the vowels /ɑː/, /æ/, and /ɒ/ for six artists representing different music genres and accents.*

7.4.2 LOT

Since there was doubt about the reliability of estimating F1 and F2 in singing, especially having used Spleeter, it was decided that it was best to avoid F3 estimations for the degree of lip rounding altogether. Another way of estimating the degree of rounding of the LOT vowel in the corpus was to compare it to data already collected for TRAP. According to Ferragne and Pellegrino (2010), /ɒ/ in BrE is more closed than /ɑː/. An unrounded American-like pronunciation of LOT should therefore shift the vowel down in the vowel space. Though it does not give a direct result of the degree of rounding, this method still provides a solid comparative basis from which to infer British and American artists' pronunciation of /ɒ/. The average Euclidean distance in F1-F2 between LOT and TRAP as a reference for the degree of openness of the vowel was therefore computed. Results can be found in Table 7.2, where the lower the value for the distance between the two vowels, the more the unrounding of LOT is likely.

Artist	Distance (Hz)
Billy Joel	398
Bruce Springsteen	437
Iron Maiden	315
Judas Priest	463
Duran Duran	519
Phil Collins	572

Table 7.2: *Average Euclidean distance between LOT and TRAP vowels in F1-F2.*

7.4.3 Classification accuracy: formants vs. MFCC

In order to find a potentially more robust representation than formants alone for the vowels under study, TRAP and BATH vowels were grouped in a unique category to be contrasted with LOT vowels. This was done for all artists except Phil Collins where only the TRAP tokens were kept. These groups were modeled according to three sets of parameters to be compared: formants alone, MFCCs alone, and a combination of formants + MFCCs. Automatic data classification was then carried out in MATLAB (The MathWorks Inc., 2024), using each of these sets as input into a support vector machine (SVM). SVM is a machine-learning algorithm that separates data points into categories by finding the best dividing boundary between them. In other words, it attempts to redefine the data in a way that makes the two groups as distinct as possible. The goal was to evaluate which representational space yielded the highest classification accuracy in terms of distinguishing TRAP-BATH from LOT vowels. The resulting

classification rates are presented Table 7.3, where higher values indicate a more robust representational space for capturing the vowel contrast.

Artist	Formants	MFCC	Formants + MFCC
Billy Joel	87.4	80.0	**90.1**
Bruce Springsteen	92.5	92.2	**92.9**
Iron Maiden	79.3	84.1	**85.0**
Judas Priest	**91.9**	88.7	**91.9**
Duran Duran	**89.7**	88.3	88.7
Phil Collins	**95.0**	94.1	**95.0**

Table 7.3: *Classification rates for differentiating* TRAP-BATH *from* LOT *vowels in percentages, using SVM. The most successful (highest) rates are indicated in bold.*

7.4.4 Discussion

Regarding the TRAP-BATH split, results clearly show most artists under study do not display this distinction. This was of course expected for Bruce Springsteen and Billy Joel, who, as AmE speakers, naturally do not have the split in their phonological systems. Looking at the vowel spaces of British artists in Figure 7.3, as well as the results from the statistical tests, it appears that most British artists do not show a significant difference in their realizations of BATH and TRAP vowels, suggesting to a certain degree that they may in fact have a more American-like realization of these vowels. However, this observation must be further examined. Both Simon Le Bon from Duran Duran and Rob Halford from Judas Priest are from Birmingham, a city located in the Midlands in England and traditionally considered to be located right on the isogloss (Chambers & Trudgill, 1998) distinguishing the region of England where the split is attested (to the South of the isogloss), and the region where it is not (to the North of the isogloss). Due to this, there is a lot of variation among English speakers in this area. This could very well be the reason why Duran Duran does not have this distinction at all, and why Judas Priest had only marginal significance in F2 values supposed to distinguish the two vowels. Although Iron Maiden singer Bruce Dickinson was born in the East Midlands, he did not live there for very long and the band was formed in London. It is difficult to definitively assert if this is due to his origins or some form of Americanized pronunciation when he sings, but auditory analysis of his speaking voice in interviews and the results from Chapter 5, tend to point toward the latter explanation. Results seem less clear regarding Phil Collins. The London-born singer does show a difference in realization between the vowels TRAP and BATH, but auditory analysis revealed that there was nevertheless a certain overlap between

the two vowels, despite it being less clear than for the other artists. Using vowel height and backness as acoustic proxies, and taking the results of the LOT unrounding analysis for Billy Joel and Bruce Springsteen as a baseline, it is possible to compare the British artists' realizations. Assuming that lower Euclidean distances between TRAP and LOT in F1–F2 space indicate a greater degree of unrounding, the results suggest that both traditional heavy metal bands, Iron Maiden and Judas Priest, produce a more American-like realization of LOT. This pattern contrasts with that of the British pop artists, whose average Euclidean distances are much higher, indicating a more rounded, British-like vowel. This may be due to heavy metal artists often singing at very high pitches. Rob Halford of Judas Priest is in fact known for his powerful screams, which require an open-mouth configuration, and may explain why the band's distance is even greater than American singer Billy Joel (463 Hz vs. 398 Hz, respectively).

While British artists seem to show a certain American-like pronunciation of their vowel sounds, this is expressed to various degrees. Overall, results point to the fact that the two heavy metal artists under study seem to be following the American model more closely in their pronunciation than the two pop artists. This does not necessarily align with previous results on the pronunciation of intervocalic /t/ for both genres, described in Chapter 5 and Chapter 6 respectively, where they had similar results overall. This is likely due to the smaller sample of artists used for this acoustic analysis, and also their own idiosyncratic approach to their music and careers. Duran Duran was in fact a pioneering band of the Second British Invasion, so maintaining a certain degree of their native accent or 'going local' (Beal, 2009) though in a nuanced way, might have contributed to their popularity with an American audience who "salivate[d] to hits from abroad" (Puterbaugh, 1983). As previously discussed in Chapter 6, Phil Collins is an interesting example as both the singer of a markedly English band in terms of identity and sound, Genesis, and as a solo artist inspired by the typically American soul genre.

Nevertheless, as was suggested in Chapter 3 which described an articulatory approach to the singing voice, and as will be further examined in Part III for /t/ specifically, further reasons pertaining to the constraints inherent to the singing voice may be at play in these different artists' realization of TRAP, BATH and LOT vowels. Because of such constraints, singers have to make compromises in their pronunciation in keeping with the concept of economy of speech gestures (Lindblom, 1990). Scotto Di Carlo (2007a) and Sundberg (1987) note that in singing, as the fundamental frequency rises, the second formant becomes much less differentiated across vowels, which makes confusion between vowels with similar F1 values such as /ɑː/ and /æ/ all the more likely. Moreover, buccal aperture is said to be greater in singing than in speech (a fact that is confirmed by the results of Production Experiment 1 examined in Part III).

Logically, this greater opening may hinder the amount of lip rounding necessary to realize a prototypical /ɒ/, and therefore leads singers to produce an unrounded vowel sound closer to /ɑ/. This effect may be partially pronounced for heavy metal artists who must deal with more extreme vocal conditions.

As for the complementary analyses using MFCC, two main conclusions can be drawn. Firstly, although the separability of vowels that are theoretically distinct does not reach 100%, classification scores reported in Table 7.3 remain robust despite the fact that measures were made under conditions that were far from ideal. This suggests, on one hand, that artists do maintain clear distinctions between phonemes that are adjacent in F1–F2 space, and on the other hand that these distinctions are at least partially measurable through manual formant analysis, in spite of recording variability, post-processing effects such as reverb or compression, and artifacts introduced by Spleeter. In addition, MFCCs, whether used alone or in combination with formant measurements, which theoretically should provide a more comprehensive representation of the phonemic contrasts under study do not add any meaningful improvement to the analysis. This is indicated by the minimal differences in classification accuracy across the sets of parameters in Table 7.3, where the greatest gap between MFCCs alone and formants + MFCC for Billy Joel is equal to 10.1 points, and the smallest gap between classification rates of the three sets of parameters corresponds to 0.7 between formants alone and MFCCs alone for Bruce Springsteen.

7.5 Summary and conclusion

By comparing the realization of the TRAP–BATH and LOT vowels in the sung productions of four British artists from different backgrounds and musical genres, and using two American pop artists as a baseline, we can see that American-like pronunciation features appear, to varying degrees. This confirms that the phenomenon is not restricted to consonants only such as the pronunciation of intervocalic /t/, but can also be seen through robust acoustic analysis of vowel sounds, as suggested by the USA-5 model model. Similar to what has been observed for /t/ in Chapter 5 and Chapter 6, this pattern seems to be influenced by sociocultural factors such as genre. Yet, articulatory factors that are inherent to the singing voice such as buccal aperture or changes in formants may also be a part of the explanation, referring back to Research Question 2 of this thesis, and the focus of Part III. Finally, the analysis of pre-existing recordings using Spleeter as a source separation tool demonstrates that acoustic measurements remain sufficiently reliable, even under non-optimal conditions.

8 The British perception of intervocalic /t/

This chapter focuses on the perception of the flapped allophone of intervocalic /t/ by native speakers of British English[1]. More specifically, it is meant to situate this feature in a diachronic perspective. As discussed in Chapter 4, it has been recognized for a long time as a defining characteristic of American English, especially in the 1980s, coinciding with the recordings studied in the previous chapters of Part II. Past predictions and more recent descriptions suggest that word-internal intervocalic flap /t/ is beginning to occur in British varieties as well, where it was not previously acknowledged as a systematic or prototypical variant[2] (Alderton, 2022; Wells, 1982a). What remains unclear, however, is whether the flap has retained its original distinctiveness as a marker of American English, or whether increased exposure to the American accent in England since the 1980s has led to a shift in the perception of the trait as considered purely American. In dealing with speaking instead of singing, this chapter seeks to clarify the extent, perception, and potential evolution of flapping in British speakers of different ages. Thereby I hope to provide a useful counterpoint–in speech–to the popular music artists' change from adopting an American accent to using Pop Song English as a supra-local pronunciation standard in their singing. As such, this chapter examines how this feature, traditionally associated with American English, may provide new insights into British English as a vector of identity. This is done by presenting an online identification perception experiment, conducted in combination

[1] This study was conducted with the assistance of Eric Ferrari, a Bachelor's student of linguistics from the University of Edinburgh, over the course of a month-long internship under the supervision of Emmanuel Ferragne at Université Paris Cité.

[2] Wells (1982a) notes that it can happen in "certain casual styles in British accents" (p. 250), but this remains anecdotal.

with a short survey.

8.1 Background

8.1.1 Perception of singing voice pronunciation

Against the background of the evolution from an overt American model to Pop Song English (PSE) as a norm in anglophone pop singing, Jansen (2018) investigated "attitudes toward and perception of an Americanized singing style in British pop and rock music" (p. 116). She did so in the form of an online questionnaire addressed specifically to British English native speakers, where participants were asked about their perceptions of American English both in everyday use and specifically in song. Results indicated that respondents were highly aware of the Americanized singing style, particularly in pronunciation, which was considered more natural and better suited to musical performance by respondents. Comparing pop and rock genres, pop singers were expected to conform systematically to this stylistic norm, whereas rock singers were more associated with local accents in the minds of British participants. Interestingly, local accent use was welcomed as a marker of authenticity and individuality, in keeping with studies focused on artists' production such as the Arctic Monkeys (Beal, 2009) or New Zealand pop singers (Gibson & Bell, 2012) (see Section 2.1.2 for a description of these studies and more). Another thing that was mentioned and will be explored in more detail in Part III of this thesis is the fact that many respondents intuitively believed that an American pronunciation was more suitable for singing in general as it "flows better with the music" (Jansen, 2018, p. 129). Jansen (2018)'s survey was in fact a pilot study for a much larger work on the same topic (Jansen, 2022), where fifty interviews were conducted face-to-face with both American and British respondents using real songs 'Americanized' to varying degrees as prompts for the discussion. Results confirmed what was found in the pilot study, and also underlined the observation that it is evident that within pop culture (PC), audiences see the United States as the dominant cultural influence, and American English is viewed as the most widespread and prestigious variety of English.

8.1.2 Aims and predictions

In her study on the perception of and attitudes toward singing style in British pop music summarized above, Jansen (2018)'s respondents identified the pronunciation of intervocalic /t/ as a flap as one of the main consonantal features of an American accent, in keeping with the USA-5 model (Simpson, 1999). While Jansen (2018, 2022)

focused specifically on the perception of contemporary pop and rock singing voice pronunciation, the aim of the study described in this chapter is purposefully broader.

Returning to Section 4.2 on the sociophonetic aspects of intervocalic /t/ pronunciation, scholars have predicted that flapping, an American feature, may in time spread to other varieties of English (O'Connor, 1991; Wells, 1982a). These observations were first made in the 1980s, and more recent studies have shown partial evidence in support of this observation, including Alderton (2022). Though his results showed that flapping is still the least favored realization of /t/ over both alveolar and glottal stops, they also revealed that it is increasingly considered as an optional contemporary variant, used for specific social-indexical purposes. Therefore, focusing on speech, this study proposes to examine the perceptual aspect of flap /t/, namely whether it is still considered a non-British feature, or has evolved into an allophone that is more generally accepted by British English speakers and listeners today. In other words, is there a generational difference in the perception of flaps as allophones that are not as unexpected as they used to be in British English? As answers to these questions, it is hypothesized that[3]:

Hypothesis 3 In an identification task, older British respondents are less likely to recognize a flap [ɾ] as an authentic British English variant of intervocalic /t/.

Hypothesis 4 In an identification task, younger British respondents will take more time in their responses than older participants when hearing a British English flap as they deem it a possible variant of /t/, whereas the older age group will only hear a /d/.

8.2 Methodology

8.2.1 Participants

The study was conducted online using the specialized research platform Prolific[4]. Participants were recruited through Prolific, which offers a large number of quality participants from around the world and ensures they are fairly compensated for their time. A total of 51 participants took part in the study. All were English (from England specifically, the rest of the United Kingdom and Ireland were excluded). For a diachronic and generational perspective on the results, participants were grouped into two age categories: 18 to 25 years old (n = 24, average age = 22) and 55 years old and above (n = 27, average age = 62). Gender was not included in the requirements or metadata.

[3]The way these hypotheses were arrived at is explained in more detail in Section 8.2.3

[4]https://www.prolific.com/. Last accessed: 16/09/2025.

Participants were given a compensation of $5 directly through Prolific for an estimated experiment time of 20 minutes.

8.2.2 Stimuli

To keep things as straightforward as possible for online respondents, the study took the form of an identification task. The stimuli were inspired by Braver (2014). His study focused on the perception of /t/ and /d/ flapping in American English by American speakers, where he used nonce words as stimuli to reduce respondents' reliance on lexical frequency and their own prior knowledge (Herd et al., 2010). Like Braver (2014), this study's nonce word tokens were three syllables long: the first syllable consisted of an unstressed CV context where C is a plosive and V is always /ə/; the second syllable included /i/, /e/, or /æ/ vowel sounds followed by the target consonant; the third syllable was always '-ing' to mimic the structure of a gerund and ensure the larger target environment was consistently VCV. All tokens followed this pattern, whether stimuli or distractors. Following Braver (2014), they were written in English orthography with the second syllable capitalized to indicate stress, and /ə/ transcribed as 'uh,' resulting in tokens such as *tuhPA_ing* or *puhKEE_ing* where the underscore _ represents the target consonant. 18 different combinations of such contexts were created. Each combination included two tokens pronounced with the two allophones of /t/ under study: *puhKEE[t]ing* and *puhKEE*[ɾ]*ing*, and two distractor consonants chosen for their contrast in voicing, place, or manner of articulation, for example: *puhKEE*[b]*ing* and *puhKEE*[p]*ing*. The full list of stimuli including tokens and distractors can be found in Figure 8.1.

The stimuli were recorded by two male native English speakers: one British English (BrE) (SSBE more precisely) speaker and one general American English (AmE) speaker. They were recorded reading 72 words each, 36 target words and 36 distractors, for a total of 144 words. Each of the 18 combinations of syllables described above had 4 target tokens: the nonce word with the [t] allophone spoken with a BrE accent, the nonce word with the [t] allophone spoken with an AmE accent, the nonce word with the [ɾ] allophone spoken with a British accent, and the nonce word with the [ɾ] allophone spoken with an American accent, as well as 4 distractors. Though unusual in these flap-conducive contexts (Patterson & Connine, 2001), the AmE speaker had no trouble pronouncing /t/ as a stop. The reverse situation where the BrE speaker had to realize /t/ as a flap was a lot more unnatural, but he had been made aware before the recording of what was expected, had been given time to practice, and could record the same word as many times as he liked until he was satisfied with the pronunciation. Unfortunately, due to a manipulation error, the BrE-stop condition (target tokens read

Context	Target stop	Target flap	Distractor 1	Distractor 2
puhKEE_ing	puhKEE[t]ing	puhKEE[ɾ]ing	puhKEE[p]ing	puhKEE[b]ing
puhKEH_ing	puhKEH[t]ing	puhKEH[ɾ]ing	puhKEH[k]ing	puhKEH[g]ing
puhKA_ing	puhKA[t]ing	puhKA[ɾ]ing	puhKA[tʃ]ing	puhKA[dʒ]ing
buhKEE_ing	buhKEE[t]ing	buhKEE[ɾ]ing	buhKEE[z]ing	buhKEE[ʒ]ing
buhKEH_ing	buhKEH[t]ing	buhKEH[ɾ]ing	buhKEH[s]ing	buhKEH[ʃ]ing
buhKA_ing	buhKA[t]ing	buhKA[ɾ]ing	buhKA[r]ing	buhKA[w]ing
tuhKEE_ing	tuhKEE[t]ing	tuhKEE[ɾ]ing	tuhKEE[p]ing	tuhKEE[b]ing
tuhKEH_ing	tuhKEH[t]ing	tuhKEH[ɾ]ing	tuhKEH[k]ing	tuhKEH[g]ing
tuhKA_ing	tuhKA[t]ing	tuhKA[ɾ]ing	tuhKA[tʃ]ing	tuhKA[dʒ]ing
puhPEE_ing	puhPEE[t]ing	puhPEE[ɾ]ing	puhPEE[z]ing	puhPEE[ʒ]ing
puhPEH_ing	puhPEH[t]ing	puhPEH[ɾ]ing	puhPEH[s]ing	puhPEH[ʃ]ing
puhPA_ing	puhPA[t]ing	puhPA[ɾ]ing	puhPA[r]ing	puhPA[w]ing
buhPEE_ing	buhPEE[t]ing	buhPEE[ɾ]ing	buhPEE[p]ing	buhPEE[b]ing
buhPEH_ing	buhPEH[t]ing	buhPEH[ɾ]ing	buhPEH[k]ing	buhPEH[g]ing
buhPA_ing	buhPA[t]ing	buhPA[ɾ]ing	buhPA[tʃ]ing	buhPA[dʒ]ing
tuhPEE_ing	tuhPEE[t]ing	tuhPEE[ɾ]ing	tuhPEE[z]ing	tuhPEE[ʒ]ing
tuhPEH_ing	tuhPEH[t]ing	tuhPEH[ɾ]ing	tuhPEH[s]ing	tuhPEH[ʃ]ing
tuhPA_ing	tuhPA[t]ing	tuhPA[ɾ]ing	tuhPA[r]ing	tuhPA[w]ing

Table 8.1: *72 nonce words created as stimuli for Perception Experiment 1, inspired by Braver (2014).*

in a British accent with intervocalic /t/ realized as a prototypical stop) was not included in the stimuli of the experiment itself–there is therefore no perception data for this condition, though it can be safely assumed that it would have been 'easiest' for British participants to respond to. A second run of this experiment is planned for later in order to resolve this error and make sure it did not impact the results reported here. The two speakers were recorded in a soundproof booth at Université Paris Cité with an AT2020USB+ microphone (Audio-Technica) at a sampling rate of 44 100 Hz and a quantization depth of 16 bits.

8.2.3 Procedure

PsychoPy (Open Science Tools, Nottingham, UK)[5] was used to construct the experiment online, and linked to Prolific to recruit respondents. They were directed to the welcome and instruction screen before proceeding to the main task. In part one, respondents heard one audio stimulus and were presented with two possible orthographic representations of the stimulus. They had to select one of the two orthographic options corresponding to what they heard by pressing the right or left arrow on their

[5] https://www.psychopy.org/. Last accessed: 16/09/2025.

keyboard. They had a practice round at the beginning of the experiment to familiarize themselves with the process, then moved on to the 144 trials for each of the randomized 144 stimuli (including both target tokens and distractors). Each trial was separated by a 500 millisecond blank interval, and both stimuli and orthographic options were counterbalanced. For the target tokens (realized either as [t] or [ɾ]), respondents had the choice between the corresponding nonce word spelled with an orthographic <t> or an orthographic <d>. The idea was to see if respondents correctly identified a [ɾ] as an allophone of /t/ by choosing the orthographic <t> option. In other words, across /t/ realizations in different accent conditions, were participants able to identify a flap as a /t/ in a nonce word they had never heard before, or would they assimilate it to a /d/? An example of a trial from the experiment is displayed in Figure 8.1 below.

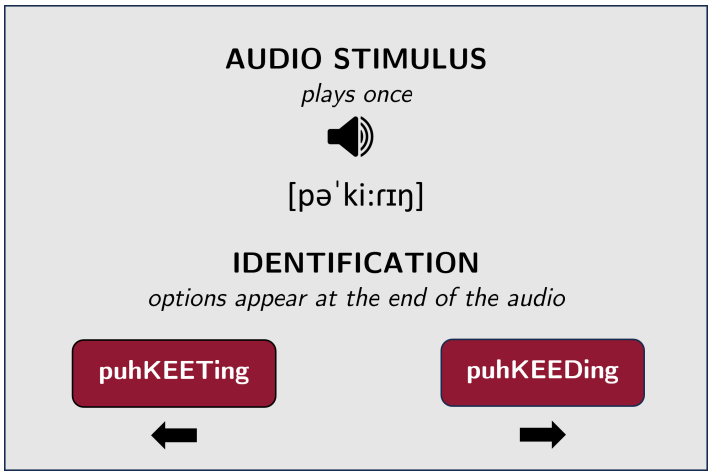


Figure 8.1: *Example trial from perception experiment. Respondents had to press the arrow key corresponding to their answer.*

In the second part of the experiment, respondents were instructed to fill out a quick survey collecting demographic information (1 and 2) as well as asking a few free response questions inspired by the ones asked in Jansen (2018)'s online experiment. A unique link for each participant was provided at the end of the experiment directing them to the survey which was built and hosted on Pavlovia Surveys (Open Science Tools, Nottingham, UK)[6]. Questions 3 to 5 dealt with attitudes to the American accent in general, and the last two were more directly aimed at singing voice pronunciation and potential media influence. Participants were required to answer all seven survey questions in an open-ended text box. The questions included in the survey were the following:

1. *What is your age?*

[6] https://pavlovia.org/. Last accessed: 16/09/2025

2. *What council/unitary authority/city/etc. are you from?* This was intended so as not to get broad answers such as "the South."

3. *Do you feel that the "American accent" has an influence on the English spoken in Britain?* [Q9 in Jansen (2018)]

4. *When and where in daily life are you confronted with the American accent?* [Q8 in Jansen (2018).]

5. *What sounds in words in American English do you think are particularly notable?* [reworded from Q6 in Jansen (2018) to focus on pronunciation specifically.]

6. *What strikes you as particularly American in the way that British singers pronounce their words when they sing?* [reworded from Q12 in Jansen (2018) to focus on pronunciation specifically.]

7. *Do you think that American media has had an effect on the way people pronounce words in British English?* This question was intended for another perspective on responses in the experiment for another study and therefore won't be examined in this thesis.

Data collection was conducted in two stages. The experiment was first opened to any potential participant matching the requirements for the younger (18-25) category. Once 25 respondents had completed the entire experiment, access was automatically closed. The experimenters manually checked to see if the respondents had completed all trials and if the time spent on the experiment was neither too short nor too long. One participant had to be excluded because of technical issues: he contacted the experimenters directly through the Prolific platform to explain that his audio stopped working. The rest of the respondents did not report any issues. Data was collected in about 40 minutes for this age category. Once everything was double checked, the experiment was re-opened to the older (55+) category, and the same process was followed. The entire set of data was collected in around 30 minutes for this age group.

8.3 Analysis

8.3.1 Preprocessing

Three types of responses were collected and analyzed in this experiment. The first one simply looked at the accuracy of the responses to the identification task. Participants' answers were coded as a match or mismatch with their orthographic transcription of the

target audio stimuli. As such, if participants heard [pəˈkiɾɪŋ] and selected puhKEETing as their answer, it was considered a match. If they selected puhKEEDing, then it was considered a mismatch. As predicted in Hypothesis 3, the idea was that older British respondents would be more often incorrect in identifying a flap as an orthographic <t>, all the more so when the token is pronounced with a BrE accent.

The second analysis, corresponding to Hypothesis 4, looked at response time (RT) as a more precise measure of allophonic perception of intervocalic /t/, rather than gross match/mismatch categorization. Per-trial response times for each participant were recorded directly in Prolific, which gave the time in seconds from the start of the keyboard component in the trial (i.e. after the audio stimuli is played, when the two orthographic choices appear on the screen) to the first key press. Firstly, outliers were identified in reaction time data, and following McLennan et al. (2005), trials with RTs more than two standard deviations from the participant's mean were discarded, which removed 2.6% of the data, leaving 3 574 trials for analysis. In addition, in order to reduce the skewness of the data, remaining RTs were log-transformed. There was another concern that older participants might be slower overall because of age-related factors and potentially less familiarity with computers. Therefore, for each participant, a baseline RT was computed from the distractor tokens to reflect average response time per participant regardless of the actual conditions under study. Then, RTs for target tokens were normalized by calculating log ratios: the log baseline was subtracted from each log RT. Resulting values are centered on zero, with positive scores indicating slower responses relative to the baseline and negative scores indicating faster responses.

The third and final analysis was qualitative in nature and consisted in collecting and identifying trends in respondents' attitudes to a potential influence of American pronunciation on British English, and identifying whether a generation gap could be inferred from their answers.

8.3.2 Statistical analysis

Statistical analysis was implemented in R (R Core Team, 2024), within RStudio (RStudio Team, 2025). The `glmer()` function from the *lme4* Bates et al. (2015) package was used for categorical outcomes (such as accuracy) and `lmer()` function was used for continuous outcomes (such as response times). Details of all models and their syntax are provided in the Results section. Plots were generated using the *ggplot2* package (Wickham et al., 2025). The p-values for fixed effects in linear mixed-effects regressions were obtained using the *lmerTest* package (Kuznetsova et al., 2017), which applies Satterhwaite's approximation to estimate denominator degrees of freedom for t-tests.

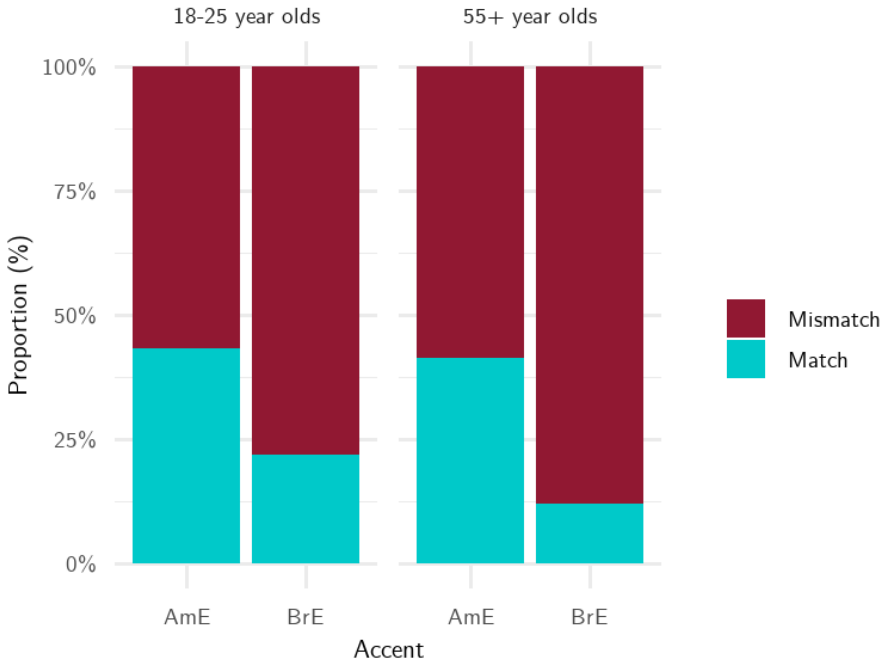


Figure 8.2: *Proportion of matches ([ɾ] correctly identified as an orthographic <t>) and mismatches ([ɾ] wrongly identified as an orthographic <d>) according to the accent of the audio stimuli and the age group of the respondents.*

8.4 Results and discussion

8.4.1 Identification accuracy

Initially, only the responses to the identification of flap audio stimuli were considered (18 flapped tokens in BrE and 18 flapped tokens in AmE per participant, for a total of 1 224 tokens). If participants correctly identified the flap as an orthographic <t> as opposed to a <d>, the answer was considered a match. The proportion of correct answers by accent and age group can be seen in Figure 8.2 below. The dark red represents a mismatch in the respondents' answer, meaning they incorrectly perceived a flap /t/ as corresponding to a <d>. Overall, the plot shows that mismatches were more common than matches for both age groups: 67% of answers were mismatches for the younger respondents and 73% for the older group. Flapping appears in general to be correctly more associated with an American English accent, with 43% of matches in AmE for the young group and 41% for the old group, against 22% and 12% for BrE tokens, respectively.

A generalized linear mixed effects model was fitted to the data including fixed effects for age groups (18-25 vs. 55+) and accent of the stimuli (BrE vs. AmE), as

well as their interaction, and a random intercept for participants. Statistical analysis was implemented in R (R Core Team, 2024), using the `glmer()` from the *lme4* package (Bates et al., 2015). While the effect of Age group alone on correctly identifying a flap as a an underlying orthographic <t> (and therefore as an allophone of /t/) was not statistically significant ($p = 0.754$), there was a significant effect of accent, where hearing a flap in a BrE accent strongly decreased the likelihood of getting a match ($p < 0.001$). The interaction of age group and accent was also significant ($p < 0.01$). Details of the model results and syntax can be found in Table 8.2.

Predictor	Estimate	Std. Error	*p* value
(Intercept)	-0.316	0.199	0.112
Age group (55+)	-0.086	0.275	0.753
Accent of stimuli (BrE)	-1.103	0.159	< .001 ***
Age Group:Accent	-0.754	0.240	< .01 **

*Match ~ Age group * Accent of stimuli + (1 | Participant)*

Table 8.2: *Output of a generalized linear mixed-effects model predicting match of flap realization to orthographic <t> as a function of respondent Generation and Accent of the audio stimuli.*

8.4.2 Response time

Response time (RT) data was also analyzed to add further precision to the identification answers by themselves. After processing the data (see description in Section 8.3 above), a three-way mixed effects linear regression model was fitted with normalized log RTs as dependent variable. Fixed effects included age group (young vs. old), audio stimuli condition (AmE-flap, AmE-stop, and BrE-flap), accuracy (match vs. mismatch) and their interaction. The model included random intercepts for participants and stimuli to account for baseline differences in overall speed across individuals and items. The detail and syntax of the model can be found in Table 8.3.

Three statistically significant effects can be identified: British flaps and American stops both lead to faster RTs than American flaps ($p < .001$ and $p < .05$, respectively). Though most of the interactions included in the model did not prove to have any effect on response time, it is nevertheless the case that when British flaps are correctly matched to /t/, responses are markedly slower than in any other condition ($p < .01$).

Predictor	Estimate	Std. Error	t value	p value
(Intercept)	−0.040	0.053	−0.757	0.451
Age group (young)	−0.041	0.076	−0.536	0.594
Accent condition (BrE-flap)	−0.101	0.024	−4.281	< .001 ***
Accent condition (AmE-stop)	−0.123	0.060	−2.044	< .05 *
Accuracy	−0.005	0.032	−0.150	0.881
GenYoung:BrE-flap	0.027	0.035	0.762	0.446
GenYoung:AmE-flap	−0.099	0.101	−0.978	0.328
GenYoung:Accuracy	−0.062	0.047	−1.312	0.189
BrE-flap:Accuracy	0.134	0.049	2.751	< .01 **
AmE-stop: Accuracy	0.003	0.067	0.052	0.959
GenYoung:BrE-flap:Accuracy	−0.089	0.066	−1.339	0.181
GenYoung:AmE-stop:Accuracy	0.130	0.110	1.178	0.239

*Response Time ∼ Age group * Accent condition * Accuracy + (1 | Participant) + (1 | Stimuli)*

Table 8.3: *Output of a linear mixed-effects regression model predicting response time as a function of age group, accent condition, and accuracy of response and their interaction.*

8.4.3 Survey

The main answers for each item (excluding demographic questions) of the survey are provided below, with questions repeated in this section for convenience. Items were designed as free-response questions, permitting participants to answer in their own words.

3. Do you feel that the "American accent" has an influence on the English spoken in Britain?

Overall, a total of 29 respondents or 60% answered yes to this question, meaning 22 (40%) said that they did not feel the 'American accent' had any influence. This proportion stays about the same when looking specifically at the two age groups, with 58% of older respondents and 56% of younger respondents answering yes. Unprompted, 15 respondents (six older and three younger ones, or about half of those who answered 'yes') mentioned they see this being the case especially for younger people.

4. When and where in daily life are you confronted with the American accent?

A majority of respondents mentioned TV as their main source of contact with the American accent, followed by internet and social media (namely video-based platforms such as YouTube or TikTok). A pie chart in Figure 8.3 illustrates the distribution of

responses according to keywords mentioned by participants.

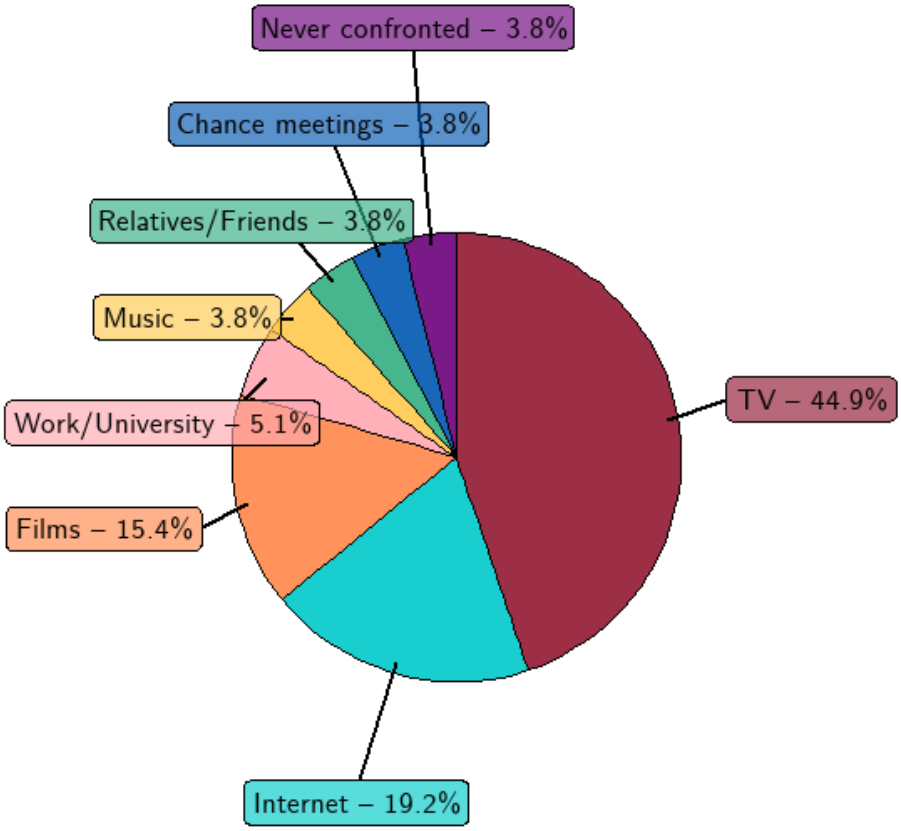


Figure 8.3: *Keywords and proportion of answers to question 4 in the survey: When and where in daily life are you confronted with the American accent?*

5. What sounds in words in American English do you think are particularly notable?

Participants' responses were adapted to phonetically accurate categories describing the main features they listed. Table 8.4 shows the count for each feature mentioned in descending order, /t/ flapping being the most mentioned by far. Even though the question was geared toward speech sounds, a few answers mentioned differences in grammar and vocabulary that are therefore not reported here.

6. What strikes you as particularly American in the way that British singers pronounce their words when they sing?

Responses to this question were more difficult to organize according to recurring themes. 18 respondents (12 from the older group and 6 from the younger group) denied hearing or noticing any use of American features by British singers. One of the rare recurring answers was diction/enunciation, though opinions were polarized: a lack of diction/enunciation was mentioned five times, but four other respondents listed exaggeration and overly articulated sounds as a tendency. Both rap and rock

Feature	Count
T flapping	10
Rhoticity	6
TRAP-BATH split	5
LOT unrounding	4
High rising intonation	2
G-dropping	2

Table 8.4: *Notable American features according to British respondents.*

were explicitly mentioned as music genres where the American accent was prevalent in British artists' pronunciation (three times and once, respectively). Section 8.5 will go over a few quotations from the survey which are particularly relevant to the topic of this thesis (singing voice), though only speech was studied in this perception experiment.

8.5 Discussion

Both the model results in Table 8.2 and visualization of the data in Figure 8.2 indicate that age group alone did not significantly affect the likelihood of identifying a flap as a match to orthographic <t>. This suggests that both younger and older listeners perform similarly in general, thus refuting Hypothesis 3. This was not entirely surprising as previous studies on flapping in American English have found that AmE speakers themselves sometimes struggle to distinguish between /d/ and /t/ flaps (Braver, 2014). Furthermore, it may have been an effect of the British speaker who recorded the stimuli sounding slightly 'off' when flapping considering it is not an allophone he would typically produce. Nevertheless, he had been trained for the purpose of the experiment, so it is assumed that this had minimal impact on the results. Turning back to the model's results, however, the strong effect of accent of the stimuli confirms previous statements that flapping is strongly associated with American English pronunciation, and much less so in a British English context. The interaction of accent and age group is interesting as well insofar as it shows that older listeners are especially unlikely to accept a flap pronunciation of /t/ in BrE contexts, while younger listeners show somewhat greater tolerance. This suggests that while both age groups associate flapping with American English first and foremost, younger British listeners might be slightly more accustomed to hearing this sound in British English, partially corroborating Alderton (2022).

Looking at the results from the response time analysis more specifically, generation did not seem to have any main effect on participants' log response time (zero-centered

by participant), as a factor by itself or its interaction with any of the other fixed effects. Both younger and older participants behaved similarly in terms of response time, and the older age group did not hesitate significantly more than the younger one when faced with flaps in BrE. However, stimuli condition did have an effect, where participants were much faster in their response time when hearing British flaps ($p<.001$), and slightly faster when hearing American stops ($p < .05$). The order of RTs from slowest to fastest is therefore: AmE-flap > BrE-flap > AmE-stop. This can be explained by the fact that AmE-stop tokens are the most straightforward in terms of stimuli, hence why respondents are the quickest when hearing these tokens. Even if the token accent is American (and with globalization and English as a *lingua franca* today, it would be difficult to assert British respondents may have a hard time understanding it), the stop /t/ pronunciation makes its association with orthographic <t> quite obvious. Furthermore, it confirms the idea that British respondents may simply be more familiar with the stop allophone. The AmE-flap condition leading to the slowest response times can be explained by the fact that participants may in fact be aware of flapping as an allophone of /t/ being a typical American trait, and therefore take longer on average to choose an orthographic representation because they are legitimately hesitating between the two. The fact that participants were significantly faster in their response time to BrE-flap audio stimuli than to AmE-flap also supports this hypothesis. This is based on the assumption that participants did not even consider the possibility of an orthographic <t> as corresponding to a flap in British English, and therefore were quick in their response to choose <d>. In other words, just because respondents selected an option more quickly for BrE-flap than AmE-flap does not necessarily mean they chose the correct answer. Looking at accuracy of responses, the interaction between a BrE-flap and accuracy is statistically significant ($p < .01$) with a positive estimate ($\beta = 0.134$): when respondents correctly match a British flap to orthographic <t>, their responses are much slower than in any other matched conditions. Thus, participants who correctly perceived a flap as an allophone of /t/ in British English were the slowest to answer. This could be interpreted as hesitation between the two options, where participants recognize that flapping is a legitimate allophone in British English, and therefore do not necessarily opt for the <d> option spontaneously.

Just as age group did not significantly affect participants' accuracy or response time, it also did not appear to influence their answers to question 3 of the survey on the influence of the American accent on British English. Within the majority of respondents who answered positively, the difference between the proportion of older and younger age groups is of two points only. Interestingly, participants from both age groups tended to spontaneously attribute this phenomenon to younger speakers. Although the data

shows no significant generational gap, participants nonetheless perceived one, identifying younger speakers as those most influenced by American pronunciation. Television was most frequently cited as the main source of this influence, consistent with previous findings that highlight its role as an important factor of language change (Stuart-Smith et al., 2013). Strikingly, the primary feature respondents state as being American is T flapping. There could well be an effect of the recurring condition in the perception experiment participants had just completed, a few of them explicitly mentioning their hesitation or confusion in the experiment related to this specific trait. This ties in with results found for response time of AmE-flap in particular, but also further supports the claim that the pronunciation of intervocalic /t/ as a flap is by far one of the main markers of American English to British people, even if its adoption in British English remains to be fully demonstrated. Lastly, question 6, though not directly tied to the hypotheses of this chapter which is focused on speech, still provides some interesting insights that bear discussing here, and relate more broadly to ideas and theories on sociolinguistic variation in pop culture and music described in Section 1.3, and discussed in the Chapter 5 and Chapter 6. A few respondents mention genre as a factor of accent shift for certain British artists, citing rock as a genre that requires "an obvious U.S. accent" or rap and hip hop as major American influences. One respondent in particular seems to have fully grasped the concept of Pop Song English (PSE) as a supra-local singing norm described by Gibson (2023): "There is nothing particularly American that I notice ... But I have noticed that quite a few artists almost seem to have a 'mid-Atlantic' accent when singing, one that is notably different to the accent they have when speaking." He recognizes there is a shift in British singers' pronunciation that is not necessarily American *per se*, but that is markedly different from how they speak. One last interesting takeaway from this survey, that links directly to face-to-face interviews carried out with four British artists and described later on in this work throughout Chapter 10 (the transcripts can be found in Appendix C.2), is the notion of a typically American lack of diction and enunciation that British singers pick up on. One respondent stated "don't sing unless you want me to hear what you sing about" when mentioning this, others qualified it as "lazy" pronunciation making "the American accent better for singing," or highlighting that "you can talk quicker in an American accent, so singers usually use it." At the basis of these comments therefore lies the idea that, in addition to genre consideration, the American accent is perceived by many British people as easier to sing, partly because of its softened or less articulate nature. Likely not being familiar with the 'Hyper-' and 'Hypo-' articulation Theory, respondents are aware of further articulatory implications and constraints of the singing voice, discussed in Section 3.3 and at the center of Part III of this thesis.

8.6 Summary and conclusion

This chapter provided a diachronic account on the perception of flapping by British respondents. The question considered was whether a younger age group may associate flapping as a typically American feature as much (or as quickly) as an older age group might. It was speculated that flapping slowly becoming an acceptable allophone of /t/ in British English (Alderton, 2022) for younger people, perceptions may have evolved and what had been considered as "one of the most striking characteristics of American pronunciation" (Wells, 1982a, p. 248) in the 1980s may be less so today. This was examined through an online perception experiment in the form of an identification task, where participants had to match flaps and stops in nonce words with their orthographic counterpart <t>. Furthermore, the study also included a survey examining British speakers' beliefs, intended to shed light on the experiment itself and gather additional qualitative data. Though it was mainly focused on speech rather than singing, apart from one survey question, results put forward by the experiment reveal interesting attitudes and beliefs that can be transferred over from one voice condition to the other.

Taken together, these results suggest that the perception of flaps is structured primarily by sociophonetic expectations about accent (AmE vs. BrE) and not by respondent age, suggesting there has not been any relevant diachronic evolution in how flapping is considered. In the identification task, both age groups behaved similarly in terms of accuracy and response time, and in the survey, both age groups reported American influence to a similar extent. Nevertheless, there is marginal evidence that younger listeners may be more open to the idea of flaps constituting a valid British English variant. Respondents frequently describe American influence as being driven by younger generations, indicating that although the quantitative data supporting this is sparse, it is *perceived* as such.

Overall this study showed that today, much like in the 1980s, flapping is perceived as a strongly American trait and a salient feature of American influence, and its integration into British English remains limited, in spite of hints that it may grow in importance in the coming years. Including a sung condition in a subsequent run of this experiment will undoubtedly provide more insight into these results. A recent study by Gibson and Hay (2025) found that as much as Pop Song English (PSE) is the norm for singing, it is also the norm for listening. Listeners from New Zealand show a *perceptual style-shift* toward PSE when listening to stimuli set to music, and have an easier time processing an AmE accent over a New Zealand English accent in that context. It would therefore be interesting in the future to examine if that is also

the case for British listeners, and if perhaps this is where the expected generational distinction, in fact, lies.

Part II: Summary

Section II consisted of an examination of the sung productions of two 1980s British genres, from the perspective of the flapped pronunciation of word-internal intervocalic /t/ as a Labovian linguistic variable, indexing either an American identity or supra-local norms of Pop Song English. It also looked at a broader linguistic context as a way to widen the scope of the genre-specific analyses. A concise summary of the main points of each chapter follows:

Chapter 5

This chapter offered a quantitative analysis of four traditional British heavy metal bands, focusing on the realization of intervocalic /t/. Results showed that all bands used the flap variant in singing, despite its absence in speech. Reasons for this style-shift driven by audience expectations, genre norms, and notions of authenticity were discussed. While diachronic patterns were less conclusive, the findings highlight how even a culturally distinct genre like heavy metal reflects broader pronunciation trends and alignment with emerging standards in popular music.

Chapter 6

Using a similar methodology to Chapter 5, this chapter provided an analysis of a representative sample of 1980s British pop music. Results echoed those of Chapter 5 in that intervocalic /t/ flapping was prevalent in sung productions, but rare in speech, confirming its role as a sociolinguistic variable indexing alignment with, or resistance to, American-inspired norms. The results suggest that the decade marked a transition from overtly adopting American accents toward a more implicit, standardized pop singing norm PSE, with artists balancing market pressures, musical inspiration, and the desire to maintain a distinct British identity.

Chapter 7

This chapter offered an exploratory acoustic analysis of vowel production, focusing on the TRAP-BATH and LOT vowels in the sung performances of four British artists from different genres, with two American singers serving as a baseline. Results showed that American-like vowel qualities do emerge in British singing, confirming that style-shift is not limited to consonants like intervocalic /t/ but also extends to vowels. This chapter also demonstrated that reliable acoustic measurements can be obtained from pre-existing recordings using source-separation tools and conventional formant extraction methods.

Chapter 8

This chapter investigated how British listeners perceive the flapped allophone of intervocalic /t/, asking whether it is still seen as a distinctly American feature. Results from both an identification perception experiment and a survey showed no major generational differences, with participants consistently associating flapping with American English. While younger listeners seemed slightly more open to its perception as a possible allophone of BrE, the study suggests flapping remains a salient marker of American speech. Further research in musical contexts would add to this perspective.

Part III

Articulatory factors

Part III: General Introduction

SPEECH sounds are always produced in context, and as Part II has shown, sociocultural and socioindexical factors have a very strong impact on the outcome of what speakers produce. However, on a more fundamental level, speech sounds are also the result of precise and automatic physiological mechanisms described in detail in Figure 3.2 in Chapter 3. This chapter gave an articulatory account of speech in general and the singing voice more specifically. Building on this foundation, Part III focuses on the articulatory factors that may lead to a shift in pronunciation from artists' speaking voice to their singing voice. In general, this part seeks to answer the second research question posed in Section 4.6. Specifically, it asks whether, beyond sociocultural factors that have often glossed over the importance of phonetic influences, articulatory features inherent to the singing voice contribute to modifications in pronunciation. Based on the realization of intervocalic /t/ variants, is there a physiological and phonetic influence on singing voice pronunciation that might lead British singers to produce /t/ as a flap rather than a stop, even when this feature is rare or absent in their speech? In addition, I investigate whether the distinct extreme characteristics of heavy metal vocals described in Section 3.2.2 also play a role, since vocal economy is a central concern in singing?

Three different production experiments were conducted with the participation of five British popular music singers in order to answer these questions. Each chapter in this part is centered on a specific study which makes use of a different experimental method: Chapter 9, or Production Experiment 1, focuses on video lip tracking, electropalatography (EPG) is addressed as Production Experiment 2 in Chapter 10, before discussing the Magnetic Resonance Imaging (MRI) study, Production Experiment 3, in Chapter 11. The main idea and original premise of the current thesis was centered around Production Experiment 2 (EPG), which explains its importance and length in this third part. Production Experiment 1 (lip tracking) was intended as a prelimi-

nary investigation into physiological aspects of singing, which served as a foundation on which to base further hypotheses. Production Experiment 3 (MRI), on the other hand, was conceived as an additional complementary and mostly exploratory examination into the phonetics of the singing voice, using this uniquely detailed imaging method.

9 Production Experiment 1: lip tracking

This chapter describes the first production experiment conducted for this thesis, using lip tracking as its main method. It is intended as a preliminary study aiming at confirming key assumptions in view of using them as a basis for subsequent hypotheses. These resulting hypotheses are tested in the central experiment of this work, Production Experiment 2, addressed in Chapter 10. Although lip tracking was meant as a preparatory study, lip tracking (Production Experiment 1) and electropalatography (EPG) data (Production Experiment 2) were recorded simultaneously, during the same procedure. As a result, there is considerable overlap between this chapter and the next, and some methodological choices have been made with one or the other experimental goal. This chapter therefore provides a description of the main aspects of the experiment that are common to both lip tracking and EPG techniques, but will only discuss in detail the features relevant to the former.

Production Experiment 1 is intended to address Research Question 2 (Section 4.6), which seeks to determine if articulatory constraints in the singing voice contribute to the observed modification of pronunciation British singers undergo from speaking to singing. In this experiment, the focus is specifically on lip aperture, as an indicator of buccal aperture more generally. The chapter will first go over the background and relevance of measuring lip aperture in singing, and describe the different methods implemented. Aims and predictions are then provided, followed by descriptions of the methodology, analysis, and results. This experiment received approval from the Ethics Committee of Université Paris Cité (included in Appendix B.1). It also received funding from the French government managed by the National Research Agency (Agence Nationale de la Recherche) under the Investments for the Future program's (programme

d'Investissements d'avenir) BRIO grant, reference 'ANR-21-EXES-0002.'

9.1 Background

9.1.1 Mouth and lip aperture in singing

It is a well-known characteristic of speech that loudness and pitch are typically interdependent: the louder an individual speaks, the more high-pitched they usually sound– the ratio found by Gramming et al. (1988) being a half semitone increase per added decibel sound level. Classically trained singers usually learn to achieve independent control over these two phonatory parameters so as to not influence the overall aesthetic effect of singing. This is less clear for CCM singers, who may not be as systematic in their voice use (Fisher et al., 2019; Sundberg, 1987). As described in Section 3.2.2, pop singing usually has a more speech-like, neutral phonation. Heavy metal singers, on the other hand, may resort to vocal techniques such as falsetto and belting, whose main features are a more open jaw and greater mouth opening. As singing requires a more open vocal tract than speaking, it is reasonable to assume that singers tend to have bigger buccal apertures than they do when speaking. In addition, these bigger apertures are likely due to the causal relationship between higher pitches and higher intensities, empirically established in speech, as attested by voice professionals of all kinds.

The distance between articulatory targets (the extent of the movement, in other words) is crucial to the dynamics of articulation (Lindblom & Sundberg, 2007), especially for the realization of stop consonants in the context of vocal economy (Lindblom, 1983, 1990). When the mouth is opened wider, the distance between the tongue and the alveolar ridge (the articulators involved in the production of a prototypical stop /t/) increases, and may impede its canonical realization. Therefore, a potential link between the distance from the tongue to the alveolar ridge as a result of buccal aperture, higher pitches and intensities, and the realization of sung intervocalic /t/ is worth exploring. Buccal aperture is the result of a complex variety of parameters, including jaw mechanics and the activity of certain muscles. For the sake of convenience, Production Experiment 1 focuses on lip aperture only, as a direct correlate of buccal aperture (Mooshammer et al., 2007).

9.1.2 Aims and predictions

The aim of this experiment is two-fold. Firstly, it seeks to verify that there is a difference in terms of lip aperture between two voice conditions: speaking and singing.

Secondly, it attempts to confirm the long-standing empirical observation made by phoneticians and singers alike that singing at higher pitches and higher intensities results in greater buccal aperture, quantified in this study through vertical lip aperture as a proxy measurement. This study was designed to fill a gap in the scientific literature on this topic, and provide solid grounds on which to base subsequent hypotheses for Production Experiment 2. Therefore, it is hypothesized that:

Hypothesis 5 Lip aperture is greater by 0.5 cm on average in singing than in speech.

Hypothesis 6 Much like in speech, lip aperture in singing increases as fundamental frequency and intensity increase.

9.2 Methodology

9.2.1 Methods for lip tracking

Lip tracking in phonetic research is typically carried out by analyzing video recordings of the participant's face. Due to its cheapness and easy implementation, it is often used as a supplementary technique in experiments involving other methods (Kochetov, 2020). Standard video frame rates (between 24 and 30 fps, or frames per second) are usually sufficient to provide information about movements of the lips and mouth (Gick et al., 2013).

Two different methods for lip tracking were used in this work. Despite pilot experiments where the first method was tested and proved effective, it had a certain number of drawbacks which influenced the decision to switch to a different method. This subsection describes both methods in detail, while Section 9.2.5 and Section 9.3 will go over how they were applied to this lip tracking study.

Method 1: Color segmentation

The first lip tracking method used in this work involves automatic color segmentation, and is partly inspired by Lallouache (1991). Subjects had to wear a headset with a stabilized camera pointed at their lips, which allows for participants to move their heads freely during the recording, as the headset enables the camera to move along with the head. In the implementation of this method, a special headset, the Ultrafit (Spreafico et al., 2018), was used. Originally designed for holding an ultrasound probe in place, it is lightweight and adjustable, and comes with an extra camera attachment to film the lips during speech. Since the arm meant to hold the probe was not needed

in this case, it was removed, leaving the arm which holds the camera. Green lipstick was applied to the participant's lips in order to facilitate post-hoc color segmentation. This was done on MATLAB (The MathWorks Inc., 2024) using the Color Thresholder app from the Image Processing and Computer Vision toolbox. A still frame of the video was used to determine an appropriate color space to isolate the green pixels of the lipstick from the rest of the colors in the video. The resulting binary mask was applied to all frames of the video and a bounding box was computed based on the highest and lowest pixels for vertical length and the left-most and right-most pixels for horizontal length. A detailed version of the MATLAB script can be found in Appendix B.2.

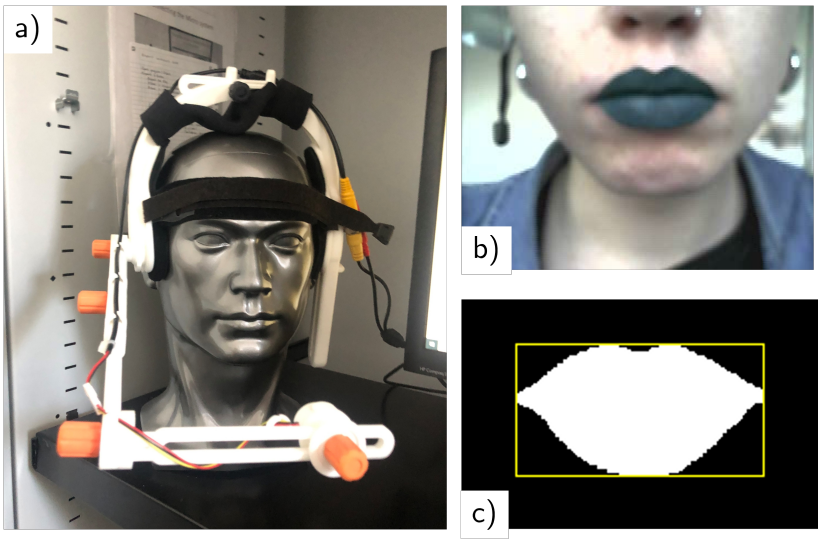


Figure 9.1: *Set up for lip tracking using automatic color segmentation, with a) the Ultrafit headset with added camera and removed ultrasound probe, b) the corresponding lip video and c) the binary mask output post-segmentation.*

Though the method was successful in the pilot tests and for the first participant (s00), it had some issues. To start with, the headset itself, although very light (350 grams), could become quite uncomfortable to wear for long periods of time. This was especially the case as it was not the only piece of equipment participants had to wear. Lip tracking data was recorded simultaneously with the electropalatography data, requiring a previously fitted palate (see Figures 10.1 showing EPG equipment). The headset was also connected by a number of cables which could hinder or disturb movements. Furthermore, although it did not impact the final data, it was difficult to obtain clear-cut color segmentation in some cases, particularly when the luminosity of the video was low–though this could be fixed afterwards in video editing software by increasing saturation. The biggest unexpected drawback to this method occurred with participant (s01), who happened to have a small patch of facial hair right between the

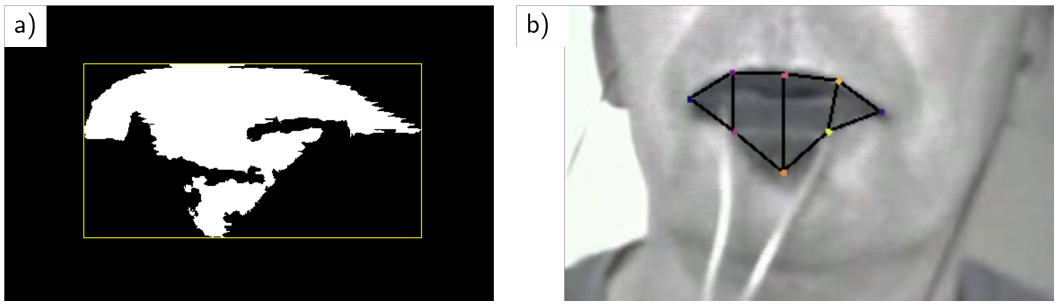


Figure 9.2: *Example of unsuccessful outputs from color lip segmentation a) and DeepLabCut lip segmentation b).*

lower lip and the chin. As a result, color segmentation was effectively impossible as no color space could reliably distinguish the lips from the surrounding facial hair. This can be seen in the binary mask in Figure 9.2.a), where the triangular area below the lips corresponds in reality to facial hair.

Another attempt at lip segmentation was done using the DeepLabCut software (Mathis et al., 2018) and a lip tracking model developed specifically for speech production (Wrench & Balch-Tomes, 2022), trained on speakers with a wide range of physical attributes (including facial hair). Unfortunately, this method was also unsuccessful at differentiating the s01's lips from his beard, as can be seen in Figure 9.2.b), with the subject's triangular facial hair being interpreted by the model as part of the lips.

Ultimately, to preserve participant s01's aperture data, a solution was found which involved measuring the length of the beard in pixels on a reference frame from the video, and subtracting it from the height dimension of the resulting bounding box for each frame. Although approximate, the method proved effective and the resulting cleaned measures credible enough that all of s01's available data was saved. However, the difficulties experienced during lip tracking by color segmentation necessitated a search for an alternative method that could avoid these issues.

Method 2: Facial landmark detection

The new lip tracking method was used on the next three participants and proved to be a significant improvement. Following Chiu and Huang (2023) and P. Krause et al. (2024), the FaceMesh face landmark detection system (Kartynnik et al., 2019) developed as part of the MediaPipe Solutions project (Lugaresi et al., 2019) was used. MediaPipe is an open-source framework developed by Google for building cross-platform multimodal machine learning pipelines, including FaceMesh. This module, implemented via Python 3.10 in Visual Studio Code (the full script can be found in Appendix B.3), detects, in real time, 468 3D facial landmarks on a human face from a single RGB image or video stream, using deep learning. One limitation is that it could not work post-hoc

on participants s00 and s01 as only the lower halves of their faces were filmed, and the module requires a full-face video in order to work. The other participants were therefore filmed using a regular webcam, without the need for additional equipment, and FaceMesh was applied to the resulting video. Figure 9.3.a shows the final result of a video frame with the mesh applied, where each point in the mesh represents a facial landmark whose relative position in the video can be retrieved. Coordinates for points 0 and 17 were retrieved according to previous lip tracking studies (Chiu & Huang, 2023; P. Krause et al., 2024), corresponding to the central points of the vermilion border of the upper and lower lips (J. C. Carey et al., 2009) (see Figure 9.3.b). These coordinates were used to compute vertical lip aperture.

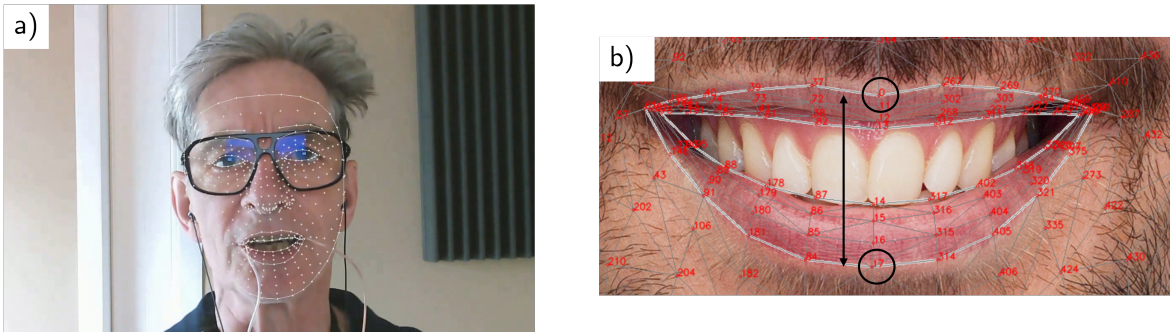


Figure 9.3: *Example of outputs from facial landmark detection a) and close-up of points 0 and 17 used to compute lip aperture b).*

9.2.2 Participants

Inclusion criteria for the experiment were quite strict in order for the data to be similar enough to allow for comparison, while remaining as close as possible to the characteristics of artists studied in the different corpora analyses described in Part II. Participants therefore had to be male, British English speakers (all were English, and an exception was made to include a Scottish participant), living in Paris, and professional or semi-professional popular music singers. A questionnaire was sent to anyone showing interest in the study, who were then vetted according to their answers. Since some of the requirements were hard to clearly define (for example, what makes a singer professional?), responses involved more than a binary yes/no answer. Regardless, a certain number of pre-established specific conditions therefore had to be met. The main criteria and how they were determined are summarized in Table 9.1. The full questionnaire including extra information can be found in Appendix B.4.

Five subjects took part in the experiment, four British professional singers and one American singer to provide a baseline. Participants were all between the ages of

Inclusion criteria	Method of determination
Male	Explicit question
British (preferably English)	Must have parents from England/UK and have been raised in England/UK
Living in France	Explicit question
(Semi-)professional singer	Must meet at least two conditions from the following: • I make (most of) my living as a singer; • I've had formal training as a singer for at least three years; • I don't have formal training but have been singing regularly for at least three years; • I perform as a singer regularly (at least once or twice a month)
Popular music singer	Explicit question

Table 9.1: *Inclusion criteria and specific method of determination for participants in production experiment.*

25 and 65, and recruited by advertising on social media, namely Facebook groups for Anglophone musicians in Paris[1] or through direct contact after extensive online and offline searches looking for British singers living in France. The flyer that was used for advertising the experiment and the information sheet provided to people who were interested can be found in Appendices B.5 and B.6, though they are more centered on the EPG experiment than lip tracking. Further discussion about the relatively small number of participants and difficulties linked to the EPG can be found in the section dedicated to participants (Section 10.2.1) in Chapter 10.

Participant s00 [control] – s00 is a 25-34 year old American professional musician from Philadelphia, currently living in Paris. He speaks English only. He described his primary styles of singing as R&B and pop. His reference high note was G^b4 and he described his accent as "American (slight Mid-Atlantic)."

Participant s01 – s01 is a 50-64 year old Scottish musician from Dundee[2], currently living in Paris. He speaks English only. Though music is not his main source of income,

[1] https://www.facebook.com/groups/378671412243699

[2] Though not from England, s01 was still included in the experiment as it was determined his /t/ pronunciation pattern resembled that of an Anglo-English accent closely enough.

and he mentioned being first and foremost a bassist, he reported performing as a singer regularly (at least once or twice a month) for at least the past three years. His primary styles of singing are pop and rock. His reference high note was G^b4 and he described his accent as "neutral Scottish."

Participant s02 – s02 is a 35-49 year old English music producer from Essex, currently living in Paris. He speaks both English and French, but sings in English. Singing is not his main activity, but he has been in the music industry for over 20 years and reported having had formal music training for at least three years, as well as performing as a singer regularly. His primary styles are rock, punk, and metal and he sings in a pop band as well as a hardcore band. His reference high note was B4 and he described his accent as "South English."

Participant s03 – s03 is a 50-64 year old English professional musician and writer with both English and French parents. He speaks both French and English but sings in English. He grew up and currently resides in England, but has lived in France for multiple years and regularly travels back. His primary styles of singing are pop and rock and he sings as a solo artist. His reference high note was A4 and he described his accent as "neutral Southern."

Participant s04 – s04 is a retired 65+ English musician from South-East London. He speaks both English and French, but sings in English. He has been living in France for over 20 years. His primary style of singing is rock, and he sings in a post-punk band, along with playing cello in a classical music trio. His reference high note was E4, and he described his accent as "lightly Cockney."

9.2.3 Stimuli

Creation

Stimuli were designed and created for use in the experiment without relying on pre-existing material such as the two corpora described in Part II. While this may have resulted in a loss of authenticity, accurately recreating the exact production conditions of 1980s recordings is impossible in an experimental setting. Moreover, this approach ensured that participants would not be influenced by prior knowledge of the songs or the way the original artists delivered the lyrics. This also made the manipulation of pitch, number of syllables and their duration, and other conditions easier.

First, a series of 14-syllable phrases were created for use as lyrics. Within the

constraints of the EPG study (described in detail Section 10.2.2). These lyrics were kept thematically as close as possible to heavy metal or more generally 1980s rock lyrics. They were set to tracks with three different melodies, whose development is described below. Four sets of lyrics and three distractors were created for each track, for a total of 21 sentences altogether. They can be found in Table 9.2. The idea behind having only three different tracks with interchangeable lyrics to learn was to reduce the individual preparation time asked of participants since they only had to learn three different melodies.

Then, three melodies were created to match the three different sets of lyrics. Using a MIDI keyboard (Arturia Keylab Essential 61 mk3), base melodies of fourteen notes (so that one note corresponds to one syllable of the fourteen-syllable sentences) were created on the Digital Audio Workstation (DAW) software Mixcraft (Acoustica, 2020). For consistency, all melodies had a tempo of 110 BPM and a time signature of 4/4, and the duration of the notes was manipulated on Mixcraft to match specific timing conditions (see the conditions paragraph below). The resulting MIDI melodies were converted with the Alter/Ego singing synthesizer VST plug-in (Plogue Art et Technologie, Inc.). It allows the user to type in lyrics corresponding to each tone, which can then be sung by a synthetic voice. A generic male voice was chosen to sing "la" for each note of the melody. This was used as a placeholder for the real lyrics, where each "la" corresponds to a syllable. Once the base melodies were created, separate instrumental tracks (drums, lead and rhythm guitars, bass) were recorded in a home studio as backing tracks to add authenticity to each melody so they would sound like excerpts from real songs. The tracks were imported in the Digital Audio Workstation and each of the three synthetic voice plus instrumental tracks were transposed from C2 to Db5, in order to have a large enough selection to accommodate to any participant's potential vocal range. Two versions of the final tracks were mixed down and exported to .mp3 format: one with the synthetic voice and the backing tracks, starting with a tempo countdown before the first measure and the second with the instrumental backing tracks only, with a tempo countdown before the first measure as well as a pitch cue (corresponding to the first note of the melody) before the countdown. The resulting duration of the final tracks is around 10 to 12 seconds. Scores for the three melodies can be found in Appendix B.7.

The final step in creating the stimuli was to produce videos that participants could follow during the experiment, using them as a guide to sing each corresponding lyric. Participants were not provided with the lyrics or these videos beforehand. MIDI tracks with the melodies were imported into Sibelius (Avid Technology, Inc., 2023), a widely used music notation software. This allowed for scores to be generated, and the 21

	Stimuli	**Distractors**
Melody 1	Grab your **lighter** and get ready we're going to **party** Smell the **metal** it's warming up, engines are **igniting** **Security** is just a myth, but my heart keeps **beating** Your **vanity** leaves me hanging but won't have me **hurting**	The energy will never fade we will not let them win In both your eyes I see the fire, you'll never surrender Just close the door and let it be, life's too short to fake it
Melody 2	Sunrise **city**, the light won't let me sleep what a **pity** All rose **petals** seem to fall to the ground in this **gutter** Don't be **bitter**, at the end of the day, nothing **matters** Write me **letters** of the beautiful world you've **created**	Voices ring out, singing all together chants of our lives The clocks tick on, reminding all of us that time is gone Alone again, around me there's only ghosts and demons
Melody 3	If it's **better** for you, then I guess I'll see you **later** She's much **sweeter** than you but they all say she's a **cheater** Look through **photos** of us, and I wonder how it **started** We've been **fighting** head on as the world around us **shatters**	There's a flame in my soul and it's burning out of control Darkness surrounding me, no light out there that I can see Shadows are on the wall but in your arms I will stand tall

Table 9.2: *Stimuli and distractors used in Experiment 1. Words containing intervocalic /t/ tokens are indicated in bold.*

different sets of lyrics were manually added. The result was saved back to a MIDI track, which now contained information about both the notes making up the melody and the newly synchronized and embedded lyrics. These files were uploaded to the software Karaoke Video Creator (Doblon, 2009), which generated perfectly aligned lyric videos. Videos were created for both the melody tracks with the sung "la la la" (used for the listening/training round) as well as for the instrumental-only tracks (used for the singing/recording round).

Conditions

The stimuli were further manipulated according to two main conditions relevant to the electropalatography experiment: syllable duration and pitch. Although the two-level pitch condition (high and regular) was designed for the EPG experiment and not analyzed in the lip tracking experiment, its presence did guarantee that there would be sufficient pitch variation to shed light on a possible relationship with intensity and lip aperture. The conditions and how they were manipulated are described in Chapter 10, Section 10.2.2.

Display

A script was created using Psychtoolbox-3.0.19 (Brainard, 1997) running on MATLAB R2024a (The MathWorks Inc., 2024). It retrieved, randomized, and displayed the stimuli to the participants according to their reference note, who could go through each trial at their own pace. The experiment is described step by step in Section 10.2.4, a diagram of the stimuli flow is represented in Figure 9.4. The full script can be found in Appendix B.8.

9.2.4 Experimental setup and material

The experiment was conducted at Université Paris Cité, in a dedicated recording room equipped with soundproof booths and a separate control area. Two computers were used in the setup, both located in the control room: a 'display' computer, used to run the MATLAB script created to present the stimuli and connected to a screen located in the recording booth; and a 'recording' computer on which the icSpeech Professional software was run to record audio, video, and electropalatography data. The booth was set up to look as similar as possible to a commercial recording studio. A high chair was available for participants to sit on, but they could also choose to stand if they preferred. Two different microphones were used. First, a t.bone BC 500 (Thomann GmbH) on a stand, connected to a soundcard (Focusrite Scarlett 18i8 3rd Gen) plugged into the

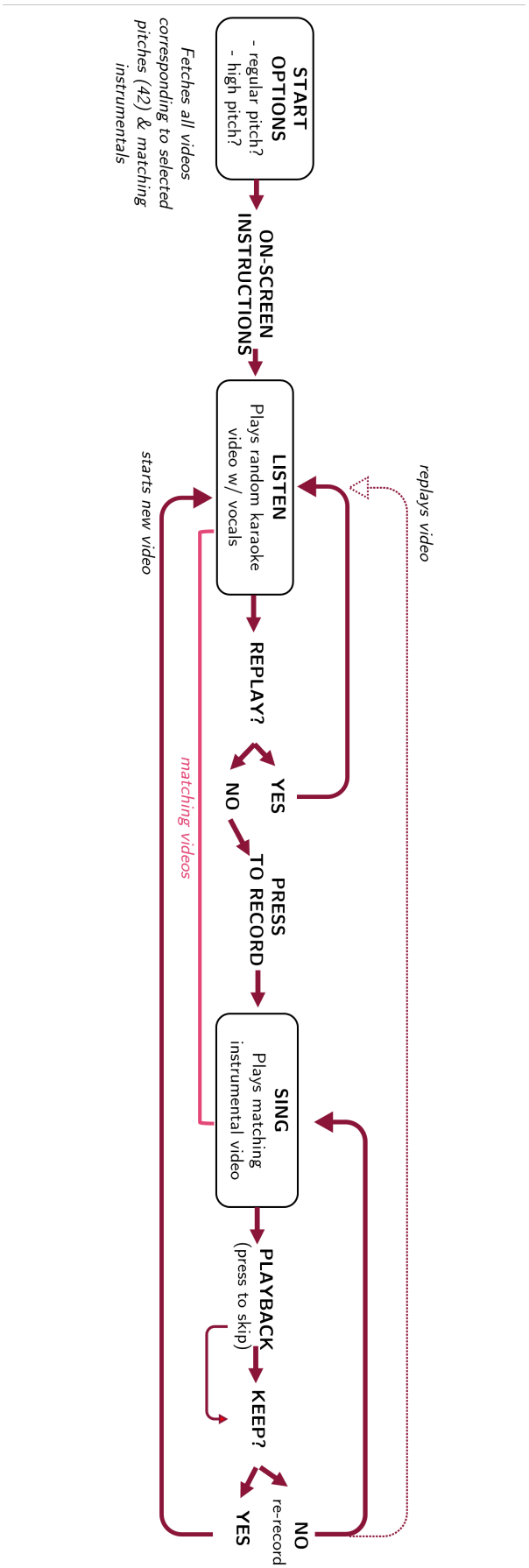


Figure 9.4: *Experimental flow for the singing part of Production Experiment 1.*

recording computer. A secondary microphone (Audio-Technica AT2020USB+) was used for live feedback only (it did not record or save any audio), attached below the first on the stand, and plugged directly in the display computer. A pair of earphones was connected to it, so participants could hear both the output sound of the stimuli and their own voice as live feedback when singing. The display screen was mounted behind the microphone stand. Participants s00 and s01, on which the color segmentation lip tracking technique was used, also wore the headset shown in Figure 9.1.a). The remaining participants did not wear the headset, instead, a webcam (Spedal AF928) was set on top of the microphone shield. Participants were provided with a wireless clicker to navigate the stimuli display. The setup for both versions of the experiment is illustrated in Appendix B.9.

9.2.5 Procedure

Throughout this chapter, only the procedure specifically pertaining to lip tracking is mentioned, though EPG data was also recorded during this process. Participants were sent audio files 10 days ahead of time to prepare before the day of the experiment. They received the four versions of all three final tracks: the one including the synthetic voice as well as the instrumental-only version (as described in Section 9.2.3), corresponding to their high-pitched reference note and transposed one octave down for the normal pitch condition. This allowed them to familiarize themselves with the three melodies they had to sing, their respective rhythm and expected pitches without being influenced by the lyrics. Lyrics were provided on the day of the experiment.

I met with the participants at Université Paris Cité. They read and signed the consent form (see Appendix B.10), and the protocol was explained to them. Blue lipstick was applied to participants s00 and s01 for the lip tracking by color segmentation method. They were then fitted with the Ultrafit headset, adjusted to their anatomy. The remainder of the participants skipped this step as color segmentation was no longer necessary for lip tracking since the facial landmark detection method was implemented.

They were guided into the recording booth where the electropalatography system was set up (see Appendix B.9). The experiment started with a practice round, where participants discovered the steps to follow for each trial and verified their reference note in case any modification was necessary. No participant required any modification. They could also manipulate the input (their own voice) and output volume (the stimuli) on the secondary feedback microphone to their own preference. The first part of the experiment involved the participant singing the randomized 42 stimuli as described in Section 9.2.3. There were two main steps involved for each trial:

1. **Listen** - First participants had to listen while the melody with the synthetic voice singing "la la la" played over the karaoke video of a randomized set of lyrics corresponding to the melody. This allowed them to discover the lyrics and practice matching them to the melody. They could replay this step as many times as they wanted. Once they felt ready, they could move on to the next step.

2. **Sing** - The same karaoke video from the first step was used. The audio began with a tempo countdown and a pitch cue for the first note of the melody, followed by the instrumental-only track. Participants had to sing the lyrics shown on screen.

After singing, participants could listen back to their recording. If they were satisfied with the recording, they could move on to the next trial. If not, they had the option of either going back to step one and listening again, or going to step two to record again straight away. This entire process is illustrated in Figure 9.4. This part lasted between 25 and 40 minutes, depending on how much time the participant spent on each trial, as everything was self-paced and participants navigated themselves through the trials using a wireless clicker. Halfway through the singing trials, participants had the option to take a break if they needed. Those wearing it were asked to keep the Ultrafit headset on, but all participants could take out their pseudopalate if they wished. s01 is the only participant who took a break and removed his pseudopalate during the experiment.

The second part of the experiment, which did not last more than 5 minutes, involved participants reading out all 21 sentences they previously had to sing, so as to have a comparison with their speaking voices possible in the analysis.

All data was recorded using the icSpeech Professional edition software provided with the EPG (see the recording computer in Appendix B.9 for a screenshot of the software). The audio was recorded at a sampling rate of 44 100 Hz and a quantization depth of 16 bits. Video was recorded at a frame rate of circa 30 fps and 720p resolution.

9.3 Analysis

9.3.1 Preprocessing

The audio and video from the resulting icSpeech file (where audio, video and EPG data are all mixed in a .wmv format) were extracted in .wav and .mp4 format respectively. Each audio file was imported in Praat (Boersma & Weenink, 2025). Both singing and speaking files were manually segmented at the sentence level in a dedicated tier. Distractor sentences were included in the analysis, for a total of 42 sung tokens (21

Participant	Spoken	Expected spoken	Sung	Expected sung
s00	21	**21**	42	**42**
s01	21	**21**	21	**42**
s02	20	**21**	35	**42**
s03	21	**21**	42	**42**
s04	NA	**21**	34	**42**
Total	**83**	**105**	**174**	**210**

Table 9.3: *Number of tokens available for analysis per participant for Experiment 1.*

sentences sung twice) and 21 spoken tokens. Not all participants' data is complete. To start with, some sung tokens were missed by the participants in the production phase of the experiment. This happened when the participant skipped to the next trial too fast and did not have time to record himself. When this happened, an attempt was made to recover another version of the token: during the first step of each trial where participants only had to listen, they were more often than not singing along already. As such, most tokens were produced twice in each trial (one version during the LISTEN step, which had not been expected, and one version during the SING step, as shown in Figure 9.4), and if one was missed, the other could potentially replace it. If the participant sang under his breath or not at all during the first step, then it was considered a missing token, and therefore no corresponding video clip could be analyzed. Participants s01 and s03 missed two tokens each this way, but one was salvaged for s01 and two for s03. The biggest loss of data was due to a software issue where the audio was correctly recorded for participant s01, but not the corresponding video. As such, 20 sung tokens were unrecoverable, and no lip aperture measurements could be extracted. Memory allocation issues caused by the long and multidimensional nature of the recording also occurred for participants s02 and s04– the audio recording cut off at specific points of the session, resulting in seven sung tokens and one spoken token missing for s02 and eight sung tokens missing for s04. Due to the same issue, all spoken tokens for participant s04 were lost, and he thus had to be excluded from the comparison between speaking and singing lip aperture. Out of a total of 42 expected sung tokens, only s00 and s03 had a complete set. There were 21 for s01, 35 for s02, and 34 for s04.

9.3.2 Extraction

For the sung audio, intensity measures for the whole file were extracted using the basic Praat option `To intensity...` where the pitch floor was set according to the

participant's estimated vocal range (with a margin of 50 Hz below their lowest pitch), the time step was set to 1 millisecond, and the option to subtract mean was activated to erase any microphone-related DC offset. The resulting intensity object was then converted using `Down to IntensityTier` → `Down to TableOfReal` and saved as a .txt file. Similarly, pitch was extracted using `Analyze periodicity` → `To pitch...` where the time step was set to 1 millisecond. Pitch floor and ceiling were set according to the participant's estimated vocal range (±50 Hz). The resulting pitch object was converted using `Down to PitchTier` → `Down to TableOfReal` and saved as a .txt file.

Lip aperture measurements were retrieved using color segmentation for s00 and s01 and facial landmark detection for the three other participants. Regardless of the method, both yielded tables of maximum vertical lip aperture measurements per video frame for each token. In other words, a sentence that took around 8 seconds to sing and recorded at a frame rate of 30 fps would have 240 individual vertical lip aperture measures. These were all recorded in pixels, which had to be converted to centimeters. In order to do so, the length in cm of each participant's right front tooth was measured on their respective 3D upper mouth models that had been made for the creation of their EPG pseudopalates (see Figure 10.6.b). The length of the same front tooth was measured in pixels in MATLAB using the Image Processing and Computer Vision Toolbox, and cross multiplied to get the pixel to centimeter equivalence for each participant. Sung and spoken lip aperture data were combined in a table, imported in R (R Core Team, 2024).

For the singing-only analysis, the participant's three tables (intensity, pitch, sung lip aperture) were imported into MATLAB as timetables[3] and synchronized together to the nearest time step for each sentence. The final table combining all participants' sung data was imported in R (R Core Team, 2024). The raw pitch data (f0) was smoothed using a simple moving average (SMA) with a window size of 3. To allow for meaningful comparisons across the different variables and participants, examined in Section 9.4.2, the raw data for intensity (dB), lip aperture (cm), and the smoothed pitch data (f0) values were standardized (z-scored) by subtracting the mean and dividing by the standard deviation, resulting in variables with mean = 0 and standard deviation = 1.

9.3.3 Statistical analysis

Statistical analysis was implemented in R (R Core Team, 2024) within RStudio (RStudio Team, 2025). The `lmer()` function of the *lme4* package was used (Bates et al.,

[3]In MATLAB, timetables are a type of table where each row is indexed by a date or time, allowing for easier handling of time-based data.

2015) to fit a linear mixed-effects models. The details of each model and their syntax are provided in the results section below. All plots were generated using the *ggplot2* package (Wickham et al., 2025). The p-values for fixed effects were obtained using the *lmerTest* package (Kuznetsova et al., 2017), which applies Satterthwaite's approximation to estimate denominator degrees of freedom for t-tests.

9.4 Results and Discussion

9.4.1 Speaking versus singing comparison

To compare the effect of voice condition (singing versus speaking) on lip aperture, a linear mixed-effects regression model was fitted in R, with lip aperture (z-scored) as the dependent variable and voice condition as the predictor. Random intercepts were included for participants and sentence. As hypothesized, voice condition was found to have a highly significant effect on lip aperture ($p < 0.001$), confirming that lip aperture considerably increases when singing as opposed to speaking. The model output is presented in Table 9.4.

Predictor	Estimate	Std. Error	t value	p value
(Intercept)	-0.358	0.055	-6.474	< 0.001 ***
Voice Condition (sung)	0.480	0.056	8.497	$< .001$ ***

Lip Aperture (z-scored) ~ Voice Condition + (1 | Participant) + (1 | Sentence)

Table 9.4: *Output of a linear mixed-effects regression model predicting lip aperture (z-scored) as a function of voice condition.*

Violin plots in Figure 9.5 show the distribution of lip aperture (in cm) in sung and spoken conditions across the four singers. As suggested by the model, the singing condition (in dark red) does indeed consistently show higher average lip aperture than the speech condition (cyan). This is reflected in the higher mean dots for each participant on the plot, and the overall higher distribution of sung versus spoken measures. Singing also tends to show a greater range of possible lip apertures across all participants, likely due to the more dynamic aspect of singing in terms of pitch and intensity, as the results for Hypothesis 6 discussed in Section 9.4.2 demonstrate. On the other hand, the spoken condition shows tighter distribution indicating more consistent lip aperture in speech. Looking at participants individually, distinct patterns tend to vary, suggesting both physical differences and a variety of vocal strategies.

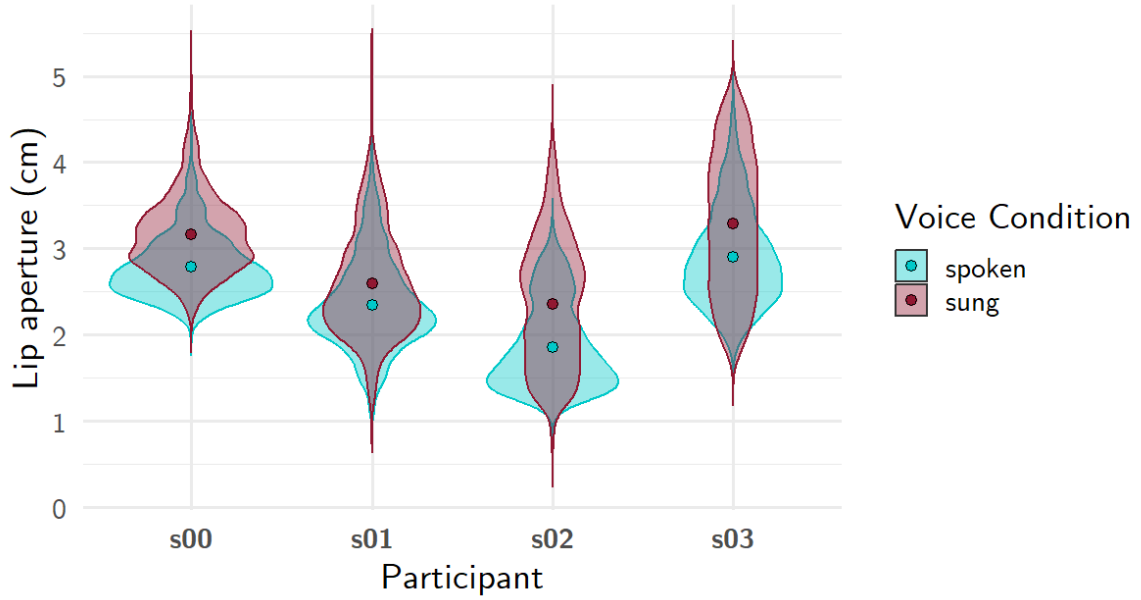


Figure 9.5: *Lip aperture distribution according to voice condition (sung vs. spoken) across four singers.*

9.4.2 Singing only analysis

Overall results

A linear mixed-effects regression model was fitted in R, with lip aperture as the dependent variable, and fundamental frequency and intensity as predictors, and their interaction. Random intercepts were included for sentence tokens and participants. Intensity was found to have a highly significant positive effect on lip aperture ($p < 0.001$), showing that higher intensity was associated with increased lip aperture. Although the main effect of pitch did not reach significance ($p = 0.806$), it is interesting to note that the interaction between the two predictors intensity and pitch did ($p < 0.01$), suggesting that the effect of intensity on lip aperture does depend on pitch level. The final model output and syntax is presented in Table 9.5.

Predictor	Estimate	Std. Error	*t* value	*p* value
(Intercept)	1.946	0.247	7.861	$< .001$ ***
Pitch	−0.00008	0.0003	−0.245	0.806
Intensity	0.008	0.001	7.780	$< .001$ ***
Pitch:Intensity	0.00001	0.000005	2.802	$< .01$ **

*Lip Aperture ~ Pitch * Intensity + (1 | Sentence) + (1 | Participant)*

Table 9.5: *Output of a linear mixed-effects regression model predicting lip aperture as a function of pitch and intensity.*

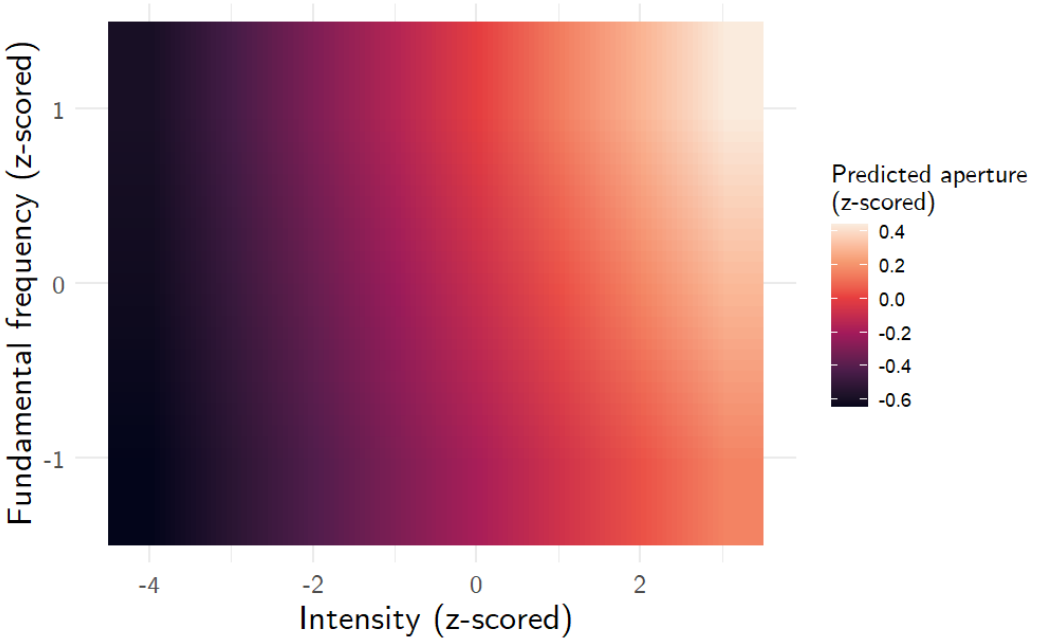


Figure 9.6: *Predicted lip aperture heatmap based on z-scored pitch and intensity interaction.*

A partial dependence plot was created to visualize the effects of pitch and intensity on lip aperture as a heatmap, shown in Figure 9.6 below. Areas of dark purple indicate regions where the predicted lip aperture is minimal, and as the color transitions to lighter shades, the predicted lip aperture increases. As the model suggests, increasing intensity (on the x-axis) leads to a noticeable increase in predicted lip aperture values. While this is not the case for pitch (y-axis) on its own at lower intensities, its influence becomes noticeably more pronounced at higher intensity levels.

Participant by participant

Since the number of participants was restricted, looking at individual participants' results can also provide an interesting view into the subtle differences that occur on an individual basis. Figure 9.7 shows the average z-scored intensity, pitch and lip aperture trends over time, the latter representing the duration of the sung melodies.

Across participants, intensity (orange) is generally the most consistently expressed parameter, with values clustering near the mean but showing smooth variation around zero. Pitch (cyan) shows greater between-participant variability, with s02 standing out as having consistently elevated pitch values compared to others. Lip aperture (dark red) appears to follow intensity relatively closely, and more so than pitch across most participants, confirming the model summarized in Table 9.5 that lip aperture is strongly influenced by intensity. Interestingly, participant s04 displays more compressed and

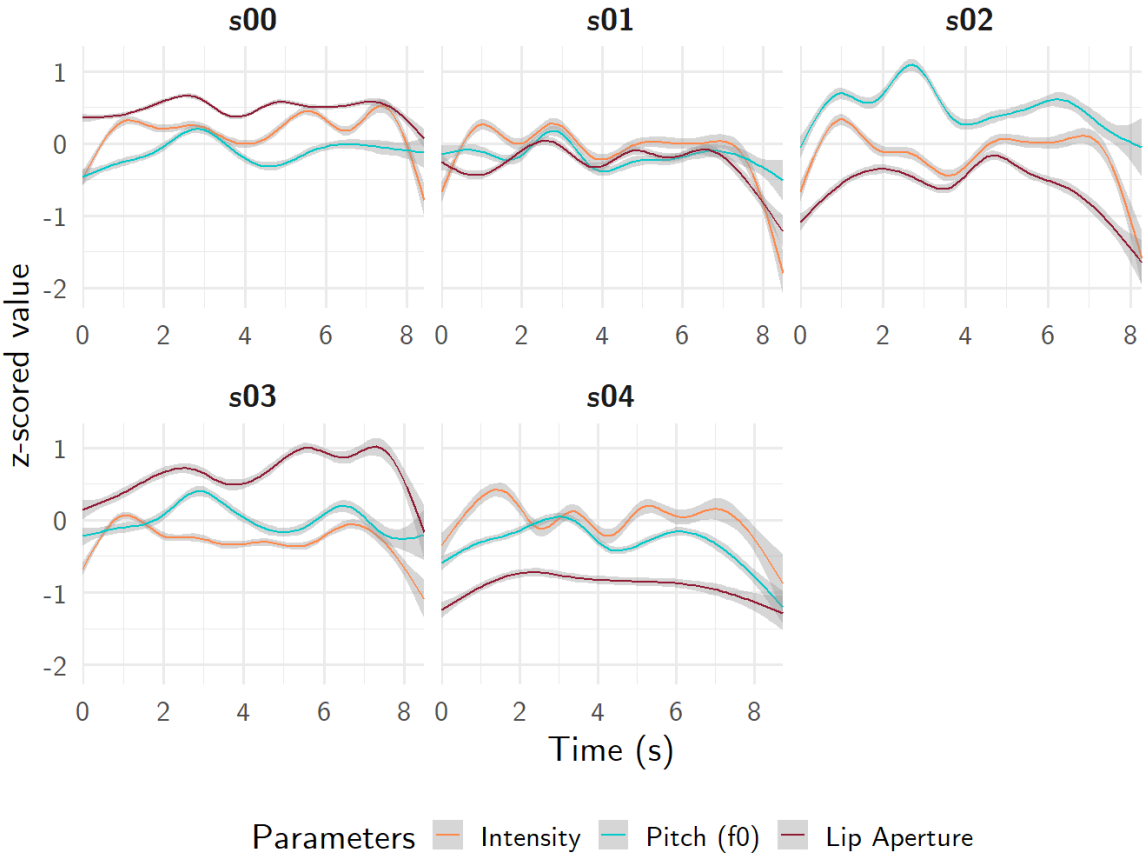


Figure 9.7: *Individual z-scored intensity, pitch, and lip aperture trends over melody duration according to participant.*

stable lip trajectory, which may suggest a more restrained performance on the day of the experiment with less exaggerated lip opening. This is interesting to put in relation with his results from Production Experiment 3 in which he also participated, discussed in Section 11.4.3, where he shows more notable variation in jaw opening. Generally speaking, participant trends confirm the existence of a relationship between the different parameters, as discussed above, but also demonstrate individual differences in their patterns. This may reflect idiosyncratic strategies in vocal performance or attitudes to the experimental environment that will have to be taken into account for the subsequent results discussed in this thesis, namely in Chapter 10.

9.4.3 Discussion

The results from the speaking versus singing comparison provide strong evidence supporting Hypothesis 5: lip aperture increases in singing compared to speaking. This is in keeping with the idea that regular speech very rarely pushes the phonatory system to its limits (Lindblom, 1990), whereas singing leads to increased effort relating to articulatory gestures (such as lip aperture) among other parameters.

The singing-only analysis attempted to take this observation further by highlighting which underlying mechanisms inherent to the singing voice may influence lip aperture, specifically looking at pitch and intensity. While intensity had a strong and significant effect on lip aperture, pitch did not reach significance on its own. However, the interaction between the two factors did prove to be significant, underlining the complexity of vocal production and the parameters governing singing voice articulation. Hypothesis 6 is therefore at least partly confirmed: lip aperture in singing increases with intensity, and, to a certain extent, with pitch at higher intensity levels.

Despite clear results, participant-level trends also highlight important individual differences in vocal strategy and articulatory behavior, which are likely due to idiosyncratic physical or stylistic features. This is important to keep in mind for further experiments, where more complex parameters may follow very different patterns. The small number of participants is not necessarily enough to compensate for such differences, especially in statistical models, but allows for a more in-depth look at personal patterns and strategies.

Broadly speaking, these results are in agreement with anecdotal and empirical observations made by voice professionals, vocal coaches, and researchers alike. Kayes (2019) and Sundberg (1987), to name two such researchers, claim that there is a link in singing between pitch, intensity and buccal aperture. This study provides quantitative support to these claims and lays the groundwork for the other studies developed in this thesis.

9.5 Chapter summary and conclusion

This chapter provided a description of the first production experiment carried out in this thesis, using lip tracking as a research method. It discussed the relevance of studying lip aperture to the research questions addressed in this thesis, and described the two main lip tracking techniques that were used, color segmentation and facial landmark detection. The methodology was provided, some of which overlaps with the following chapter as both experiments were carried out simultaneously. Participants, stimuli, materials and procedure were delineated. A two-step analysis was conducted, firstly comparing lip aperture measurements between speaking and singing, before focusing on singing-only and investigating pitch and intensity as parameters of influence on lip aperture.

The results mostly conform to the hypotheses set out in Section 9.1.2, which were posited as a starting point for subsequent analyses looking at articulatory patterns in the singing voice. Specifically, it is of interest whether increased lip aperture may lead to pronunciation strategies in keeping with vocal economy as discussed by Lindblom (1983). This is done using EPG and MRI as additional experimental methods (see Chapter 10 and Chapter 11, respectively).

10 Production Experiment 2: Electropalatography (EPG)

This chapter describes the second production experiment conducted for this thesis, using electropalatography (EPG) as its main method. This study is the central experiment of this thesis, framed by the preliminary Production Experiment 1 addressed in Chapter 9 and the exploratory Production Experiment 3 addressed in Chapter 11. It is therefore considerably longer and more detailed. This experiment was conducted at the same time and using the same protocol as Production Experiment 1 on lip tracking. As a result, the methodology is identical and only new elements pertaining to the present study will be discussed in depth, though references to previous sections will be made throughout.

The experiment presented in this chapter seeks to answer Research Question 2, specifically by looking at linguopalatal contact in the singing voice as a way to identify and observe factors inherent to singing which may contribute to variation in intervocalic /t/ outcome, either as an auditorily perceptible flap, or at the very least in terms of contact patterns of the tongue against the palate. The chapter begins with an introduction to electropalatography, providing background on the method and describing the aims and predictions of the present experiment. Methodology and the various post-hoc data processing and analyses are outlined. Results for each hypothesis and their respective discussions are provided in Section 10.4, before the final summary and conclusion of the chapter. This experiment received approval from the Ethics Committee of Université Paris Cité (included in Appendix B.1). It also received funding from the French government managed by the National Research Agency (Agence Nationale de la Recherche) under the Investments for the Future program's (programme d'Investissements d'avenir) BRIO grant, reference 'ANR-21-EXES-0002.'

10.1 Background

10.1.1 What is electropalatography?

Electropalatography is a technique that allows for the real-time identification, recording, and display of the points of contact of the tongue on the hard palate during speech production (Hardcastle, Gibbon, & Jones, 1991). It consists of a personalized pseudopalate made of a thin plate (similar to a dental retainer) in which electrodes are embedded. The electrodes detect lingual contact against the pseudopalate during articulation and transmit the signal through wires that exit in two bundles from the corner of the participant's mouth. The signal is then processed by the external electropalatography (EPG) unit (also called a multiplexer), usually worn around the neck, and transmitted to a connected computer. A bracelet or hand grip also linked to the multiplexer supplies a small current. A conductive path is formed every time the participant's tongue touches the electrodes and real-time display of the areas of contact is shown directly on the computer screen on a dedicated software program. The resulting display usually takes the shape of a grid or matrix where one square corresponds to one electrode or, alternatively, appears on an anatomically correct occlusal diagram of the palate. The instrument and display are shown in Figure 10.1, which depicts, from left to right, the display screen (anatomical view) and the external unit connected by wires to the pseudopalate in the participant's mouth.

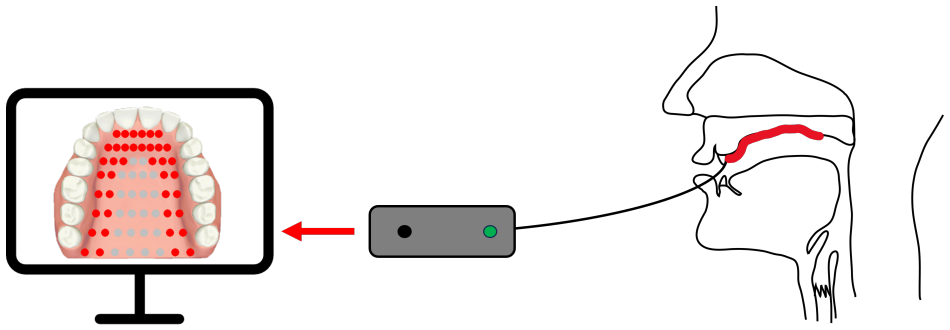


Figure 10.1: *Diagram of an electropalatography (EPG) system. The acrylic pseudopalate (in red) in the participant's mouth is connected to the multiplexer (in gray), plugged into a computer displaying real-time contact. The anatomical diagram of the palate is taken from the icSpeech Professional software by Rose Medical Solutions Ltd.*

The current version of electropalatography has evolved from older techniques of static palatography, which allowed the recording of a single location of tongue-palate contact for an isolated sound only. Palatography as a research technique in speech and articulation was developed in the 1950s and first described by Ladefoged (1957) and

Abercrombie (1957). It consisted of painting the tongue with a mixture of charcoal and other substances like vegetable oil or chocolate powder. The speaker then produced a single occurrence of a sound and a photograph of the palate was taken to study the areas where the mixture rubbed off on contact. Figure 10.2 shows an example of static palatography taken from Ladefoged (1957), the black stains on the palate indicate the places where the tongue constriction happened during the articulation of /k/ before front and back vowels[1]. First attempts at resolving the limitations of static palatography–namely, that it can only identify one constriction pattern at time before the articulators have to be cleaned and the mixture reapplied–were described by Kydd and Belt (1964) and Shibata (1968), focusing on the observation of dynamic changes in tongue-palate contact.

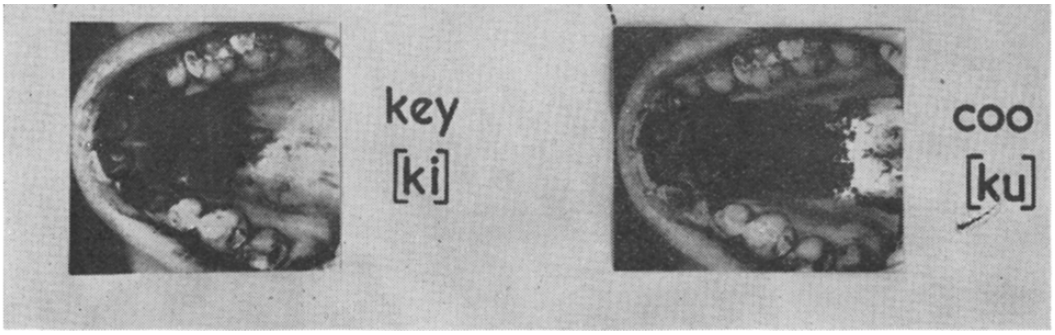


FIGURE 3. Palatograms of *key* [ki] and *coo* [ku].

Figure 10.2: *Example of static palatography taken from Ladefoged (1957), illustrating the articulatory differences of /k/ when followed by a front and back vowel.*

There are currently four versions of EPG systems commercially available: the SmartPalate system (CompleteSpeech, USA), the WinEPG (Articulate Instruments Ltd., UK), the WinSTARS (EPG Research Centre, Japan), and the LinguaGraph (icSpeech, Rose Medical Solutions Ltd., UK). They all follow the same general principles but differ in terms of palate shape and fit, number of electrodes, and temporal resolution. I will be referring exclusively in the rest of this work to the LinguaGraph system developed by icSpeech as it is the one used for the current experiment (see Section 10.1.2 for a more detailed description), but see both Lee et al. (2022) and Mat Zin et al. (2021) for comprehensive reviews of the different EPG systems mentioned above.

There are some significant advantages to using EPG as a linguistic research tool. Its high temporal resolution (200 Hz for the LinguaGraph) and precise electrode placement allow it to record in real-time fine-grained detail about timing and location of tongue placement during continuous speech. It is also one of the only visualization techniques

[1]Incidentally, a set of pedagogical resources on the allophones of /k/ in French that I compiled using EPG and MRI can be found here: https://github.com/CCaillol/Kuki

that provides precise information on lingual contact against the whole palate, and from a non-sagittal view[2]. Nevertheless, there are some significant limitations as well. EPG can be very costly and time-consuming as each participant must have a custom-made pseudopalate which can amount up to 700€ a piece and takes on average a month to produce. For this reason, based on a survey of 379 full research articles, Kochetov (2020) found that the median sample size for EPG studies was four participants, much lower than other comparable methods. Furthermore, it does not provide any information such as the shape of the tongue or its location when it is not touching the palate, and spatial resolution is limited to the number of electrodes embedded into the palate (ranging from 62 to 124 depending on the model), most of which miss dental or very retracted velar articulations. Nevertheless, it has been shown that more electrodes do not necessarily equate to better results as it may make single contacts harder to distinguish, among other things (Fougeron et al., 2000; Wrench, 2007). Another thing to take into account is the fact that, although the pseudopalate is custom-made to fit each participant, it is nevertheless an unfamiliar device positioned in their mouth which may cause discomfort, excess salivation and speech distortion, which crucially may impact the validity and generalization of the results obtained. It is customary to wait a certain amount of time after first wearing the palate so that the participant can adapt. In a majority of cases, 45 minutes seems enough for speakers to get used to their pseudopalates, though it seems that some cannot adapt at all (McAuliffe et al., 2008). In general, the articulation of fricatives is the most affected, though in one study, a panel of expert listeners indicated little to no change in the speech naturalness of speakers wearing a pseudopalate after adaptation time (McLeod & Searl, 2006). More importantly, it seems that the least affected phoneme is /t/, which appears to be unaltered perceptually and spectrally (McAuliffe et al., 2008). Despite several drawbacks which must be taken into consideration, electropalatography has imposed itself as an appropriate technique to use in this thesis to study and compare the sung articulation of intervocalic /t/ variants as lingual consonants.

10.1.2 The LinguaGraph unit

The EPG system used for this research is the LinguaGraph system developed by Rose Medical Solutions Ltd. It was created specifically for clinical purposes (Kelly et al., 2000), which also makes it ideal for research, since it is portable and both user and patient/participant friendly. The LinguaGraph uses a similar system to the Reading palate, first developed in the 1960s and updated until the 1990s (Kelly et al.,

[2]Ultrasound can provide images on the coronal plane, but the tip of the tongue is much more difficult to visualize than using EPG.

2000). Their latest EPG3 system has since been replaced by WinEPG (Wrench, 2007), mentioned in Section 10.1.1 above. The basic characteristics of the Reading system (Hardcastle, Gibbon, & Jones, 1991; Hardcastle et al., 1989) remain the same for the LinguaGraph: an acrylic artificial plate fitted with 62 silver electrodes wired to a double-sided circuit gold-plated connector, which plugs into the EPG multiplexer unit.

The 62 electrodes are arranged on the pseudopalate according to a predetermined pattern based on anatomical landmarks. This provides sufficient coverage of the majority of the palate, including velar and alveolar regions, and critically allows for the comparison of anatomically different palates as the reference positions remain similar relatively-speaking across subjects. Three reference lines are followed to help with electrode placement:

1. A horizontal line across the top at the juncture between the upper front teeth and the palate (alveolar ridge). See line 1 in Figure 10.3.a).

2. A horizontal line in the middle of the juncture between the hard and soft palate (velar region). See line 2 in Figure 10.3.a).

3. A vertical line down the center of the palate. See line 3 in Figure 10.3.a).

The electrodes are then arranged according to these landmarks in eight horizontal rows of eight electrodes, except the first row which only has six. The electrodes are spaced at equal horizontal intervals, while vertically the four front rows are arranged with half the spacing of the four back rows. The electrodes are thus further apart in the back section of the palate. The palate is therefore divided into three main regions: alveolar, palatal and velar. The individual electrode wires are bundled up in tubing and exit from each side of the mouth. They are soldered onto the connector which can then be plugged into the EPG multiplexer. The LinguaGraph system uses a recording, visualization and analysis software called icSpeech, whose interface can be seen on the recording computer in Appendix B.9.

10.1.3 Electropalatographic studies

Though it is expensive, electropalatography is often used in clinical settings as well as speech and language therapy, mainly because of the easy-to-interpret real-time biofeedback it allows. As such, EPG has been used to help patients with speech sound disorders (Gibbon & Wood, 2010; McAuliffe & Cornwell, 2008), cleft palates (Gibbon et al., 1998), Down's syndrome (Cleland et al., 2009), dysfluency (Forster & Hardcastle,

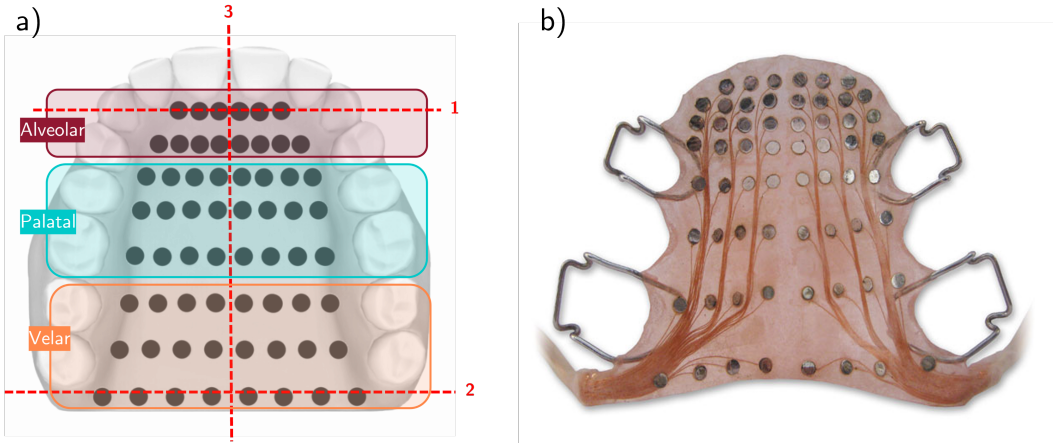


Figure 10.3: *a) Diagram of anatomical landmarks and places of articulation for the LinguaGraph pseudopalate and b) picture of a Reading-style LinguaGraph pseudopalate. Image b) taken from the icSpeech website https://icspeech.com/default.html.*

1998), hearing impairments (Ellis, 2009), and neurological disorders such as Parkinson's (McAuliffe et al., 2007). This section will only focus on linguistic and phonetic research applications, though see Gibbon (2013), Hardcastle and Gibbon (2005), and Lee et al. (2022) for reviews of EPG as a clinical and speech and language therapy tool.

EPG in articulatory phonetics

In addition to its clinical use, electropalatography is also an important tool in articulatory phonetics, on atypical and typical speech alike. Electropalatographic research in phonetics can generally be sorted into three main categories, though this is by no means an exhaustive representation of the wide variety of EPG studies: articulatory description and categorization of various consonantal sounds, mechanisms of coarticulation, and connected speech processes. Table 10.1 provides an overview of a few representative examples of such studies.

On the articulation of /t/

EPG is particularly useful to investigate the articulation of coronal consonants which are produced with the flexible front part of the tongue. This includes studies on alveolar stops and more precisely various realizations of the phoneme /t/. This subsection complements Section 4.1 on the pronunciation of intervocalic /t/ in English, looking through the lens of linguopalatal contact. Tongue-palate patterns characteristic of a prototypical alveolar stop /t/ have been looked at in a number of EPG studies, confirming articulatory descriptions found in other impressionistic or instrumental works.

Main research subject	Topic of investigation	Study
Production of coronal consonants	Korean coronal obstruents	Anderson et al. (2004)
	Velar stops in English	Liker and Gibbon (2008)
	Coronal place constrast in Argentine and Cuban Spanish	Kochetov and Colantoni (2011)
	English coronal stops	Sung and Kochetov (2018)
	Contact in Japanese geminates and singletons	Kochetov (2025)
Coarticulation	Consonant-to-vowel coarticulation	Recasens (1991)
	Lingual coarticulation in /kl/ sequences	Gibbon et al. (1993)
	F1 coarticulation in VCV sequences	Recasens and Pallarès (2000)
	Coronal consonant coarticulation in Central Arrernte	Tabain and Beare (2022)
Connected voice processes	Asymmetry of English sibilant assimilation	Pouplier et al. (2011)
	Overlap and gradient assimilation in French nasals+stop sequences	Steele et al. (2019)
	assimilation of English word-final /n/	Kochetov et al. (2021)
	Reduction of word-final /l/ in Spanish	Ramsammy (2023)
Prosody-phonetics interaction	Articulatory strengthening in prosodic domain-initial position	Fougeron and Keating (1996)
	Effects of repetition and lexical frequency	Stephenson (2003)
	Initial clusters in prosodic conditions	Bombien et al. (2006)
	Effects of initial position and prominence on initial CV	Cho and Keating (2009)

Table 10.1: *Overview of representative EPG studies.*

Most studies on /t/ consider it alongside other alveolar sounds such as /d/ and /n/. This provides a useful basis for the comparison of voiced and voiceless stops, which is part of what makes the distinction between [t] and [ɾ] (see the paragraph on voicing in Section 4.3.1).

Voiceless alveolar stops have been reported to display more linguopalatal contact than voiced alveolar stops (Fletcher, 1989; Moen & Simonsen, 1997), a result since corroborated with ultrasound data (Ahn, 2015). This has been linked to the increase in air pressure needed during the articulation of a voiceless stop, as compared to a voiced one. In other words, since stronger oral air pressure is required during the articulation of /t/ than /d/, a firmer and therefore greater contact between the tongue and the alveolar ridge is needed in order to prevent air from escaping.

Furthermore, in addition to the expected anterior constriction, Gibbon et al. (2007) found that /t/ also shows a significant amount of bilateral constriction. More recently, Lee et al. (2015) have also shown a consistent rapid increase in lateral contact during the onset of alveolar /t/, interpreted as an upward movement of the sides of the tongue right before alveolar closure. These results coincide with other findings on lateral tongue bracing, and can be seen as a secondary place of articulation required in the production of stop /t/. As a result, the prototypical stop /t/ configuration has been in turn described as "spoon-shaped" (Lee et al., 2015) or "horseshoe"-like (Gibbon et al., 2007). This can be seen clearly in Figure 10.4, which shows an EPG display of the tongue-palate pattern for a prototypical stop /t/, with activated electrodes in both the alveolar and lateral areas of the pseudopalate. Lateral tongue bracing during the production of stop /t/ plays a major role in providing overall tongue stability (Fuchs et al., 2006; M. Stone et al., 1992) and allowing a tight and complete seal to be created in order to build up enough intraoral pressure for the burst release of the plosive consonant (Fuchs et al., 2006), as described in Section 4.1 in Chapter 4. Part of the reason why EPG was selected as the main experimental tool is that lateral tongue bracing is a type of articulatory phenomenon that can only be seen directly with an EPG system, since, as opposed to most other imaging techniques, it provides a top-down occlusal view of the palate. However, this comes with the caveat that even when the EPG indicates such tongue bracing, there may nevertheless be a small amount of air escaping into the mouth that is not detectable (Gibbon et al., 2007).

In addition to studies on alveolar stop /t/, a few other works have looked more specifically at flap /t/ using EPG. Recasens (1991, 2007) investigated the realization of taps and flaps in Catalan, reporting that in part due to their very short (or even at times nonexistent) closure, the positioning of the tongue body does not require much articulatory control. In other words, as opposed to the precise tongue-palate

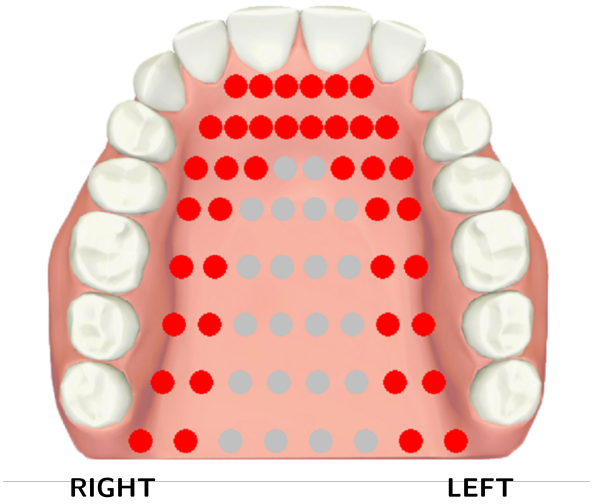


Figure 10.4: *EPG tongue-palate contact pattern for a prototypical alveolar stop /t/. The view is from an occlusal plane, an imaginary flat surface that touches the tips of the upper and lower teeth where they meet when biting. Image generated using the icSpeech Professional Edition software (Rose Medical Solutions, Ltd.).*

positioning required for the production of stop /t/, involving both complete alveolar closure and lateral bracing against the palate, flaps are associated with a relatively low degree of tongue constraint. In English, flapping has been addressed in a wider study on the positional and contextual variation in the realization of Canadian English coronal stops /t/ and /d/ (Sung & Kochetov, 2018). Sung and Kochetov (2018) looked at these in word-initial position, word-final preconsonantal position, and, as a specific context for flapping, focused on the pronunciation of /t/ across word boundaries in word-final prevocalic (V_#V) position. Though this thesis focuses on word-internal flapping as opposed to flapping across word boundaries, it may be assumed that their results apply to other flapping contexts. Final /t/ was found to be articulated with less linguopalatal contact and shorter closure duration, demonstrating systematic weakening and thereby flapping-related lenition for prevocalic positions. In other words, the articulation of /t/ in a flapping context is clearly different from word-initial /t/, where it can only be realized as a stop. The latter displays significantly more tongue-palate contact and longer closure duration than the former. Overall, existing electropalatography studies on both stop and flap /t/ have contributed to characterize and confirm many previous observations about these allophones. These studies, as well as the more general descriptions of both realizations of /t/ provided in Chapter 4, allow predictions to be made with regards to this EPG experiment, looking at the realization of intervocalic /t/ in the singing voice.

10.1.4 Aims and predictions

As outlined in the Section 4.6, the main aim for this experiment is to explore how articulatory considerations inherent to singing may also contribute to the realization of one allophone of /t/ over another, based on the pronunciation features of intervocalic /t/ variants and the principle of economy of speech gestures, previously described in Section 4.3.1. In other words, the general prediction for this experiment can be summarized as:

General hypothesis The more extreme or difficult the singing context, the more likely a flap or flap-like realization implying less linguopalatal contact of word-internal intervocalic /t/ is to occur.

Consequently, the following predictions can be made:

Hypothesis 7 There is on average less linguopalatal contact in singing than in speech.

Hypothesis 8 If singing leads to greater buccal aperture (see Experiment 1 for Hypothesis 5) and this effect increases at higher pitches and intensities (see Experiment 1 for Hypothesis 6), it is expected that, at higher pitches, word-internal intervocalic /t/ will be realized as a flap, or at least with reduced linguopalatal contact.

Hypothesis 9 Since flaps are much shorter than stops in terms of closure duration, it is expected that, at faster rates of singing, word-internal intervocalic /t/ will be realized as a flap, or at least with reduced linguopalatal contact.

10.2 Methodology

10.2.1 Participants

The inclusion process for participants was described in Section 9.2.2, which also provided a brief description of the profile of the five singers who took part in the experiment. The small number of participants[3] in this production experiment is due to different factors. Firstly, as previously mentioned in the description of the system (Section 10.1), electropalatography (EPG) can be expensive, as it requires each participant to be fitted with a custom-made pseudopalate. Additionally, very strict

[3] Although it is relatively small compared to other experimental techniques, four participants is the median sample size in electropalatography studies (Kochetov, 2020).

inclusion criteria made it difficult to find many people who met all the requirements and were interested in the study. Particularly challenging was the necessity of being a native English speaker from England and living in Paris (or at least visiting monthly). This criterion was imposed because the EPG pseudopalate-making process (described in Section 10.2.3 below) is quite long, which introduced a month-long delay between taking the participant's palate impression and the pseudopalate being ready for use. Therefore, short-term visitors or touring artists could not participate as very few of them would be available at two separate times spanning a month's interval. Similarly, since the dental lab creating the pseudopalates was based in Paris and the experimental setup quite elaborate, it would have been difficult to travel to England to recruit and record participants. All in all, it took six months to find five subjects and have them all complete the experiment successfully.

10.2.2 Stimuli

Creation

The creation process for the stimuli is described in Section 9.2.3. The lyrics (provided in Table 9.2) all feature intervocalic /t/. As can be seen in bold in the table, two words with an intervocalic /t/ that could potentially be realized as a flap were included in each sentence, on syllable four (referred to as flap 1) and syllable fourteen (referred to as flap 2). The lyrics were written according to these constraints.

Condition: pitch

Two conditions were manipulated as variables in the stimuli to match the two predictions made in Section 10.1.4. The first was related to the lip tracking experiment described in Chapter 9, specifically the results of Hypothesis 2 looking at the influence of pitch and intensity on lip aperture. Though intensity emerged as a clearer predictor than pitch on its own, it was more challenging to directly control it in the EPG experimental design. Intensity appeared more difficult to standardize, and instructing singers to pay attention to this additional parameter could have led to unnatural performances. Instead, pitch seemed relatively easy to control using melodic cues, for example. Singing in different keys or at different pitch levels is also something any trained singer can do without giving it much thought, so it would not influence the outcome. Participants therefore had to sing the 21 sentences corresponding to three melodies twice (for a total of 42 sung sentences), once at a high pitch and once at a regular pitch. The pitches were individually determined beforehand with the participants, so that it would fit each person's own vocal range. They were asked to determine

the highest note they could produce comfortably–not the highest possible note in their vocal range, but one that was high enough but that could be produced without too much strain. This note was then used as the reference note for the selection of the key for the three melodies: it had to correspond to the highest possible note that could occur in the melodies. For example, if a singer said their high note was A4, then the highest note in the three melodies had to be A4. For the regular pitch, everything was transposed one octave down, given that a vocal range of about two octaves is common for professional singers (Fisher et al., 2019). In keeping with the previous example, the highest possible note to occur in the melodies for the regular pitch condition would therefore be A3.

Condition: syllable duration

The other condition was syllable duration. Results from the duration analysis of flap and stop /t/ tokens in heavy metal and pop music productions are discussed in Chapter 5 and Chapter 6 (reproduced in Table 10.2 for convenience), and confirm that duration is an important factor of distinction between the two /t/ variants, worth including in this experimental approach.

	Flap duration (ms)	**Stop duration (ms)**
Heavy metal	592	727
Pop	542	670

Table 10.2: *Duration (in milliseconds) of intervocalic flap and sung /t/ in the sung productions studied in Chapter 5 for heavy metal and Chapter 6 for pop genres.*

Each note in the melody corresponded to a syllable in the lyrics. As it can be argued that intervocalic /t/ occurs at the boundary between two syllables, both syllables of the trochee in which the target /t/ occurs were controlled. Syllable 2 (such as *-ter* in *better*) consistently consisted of a quarter note (1/4). The duration of the first syllables were determined according to the three types of melodies and the placement of the /t/ token in the sentence (flap 1 on the fourth syllable and flap 2 on the fourteenth syllable). For melody 1, syllable 1 of flap 1 was a quarter note (1/4) and syllable 1 for flap 2 was an eighth note (1/8). For melody 2, syllable 1 of flap 1 was a half note (1/2) and syllable 1 of flap 2 was a quarter note (1/4). For melody 3, syllable 1 of flap 1 was an eighth note (1/8) and syllable 1 of flap 2 was a half note (1/2). In total, this amounts to 8 tokens for each duration of syllable 1, and 24 tokens of the same duration for syllable 2. Figure 10.5 shows a diagram of all the experimental conditions, illustrating pitch and note duration variations.

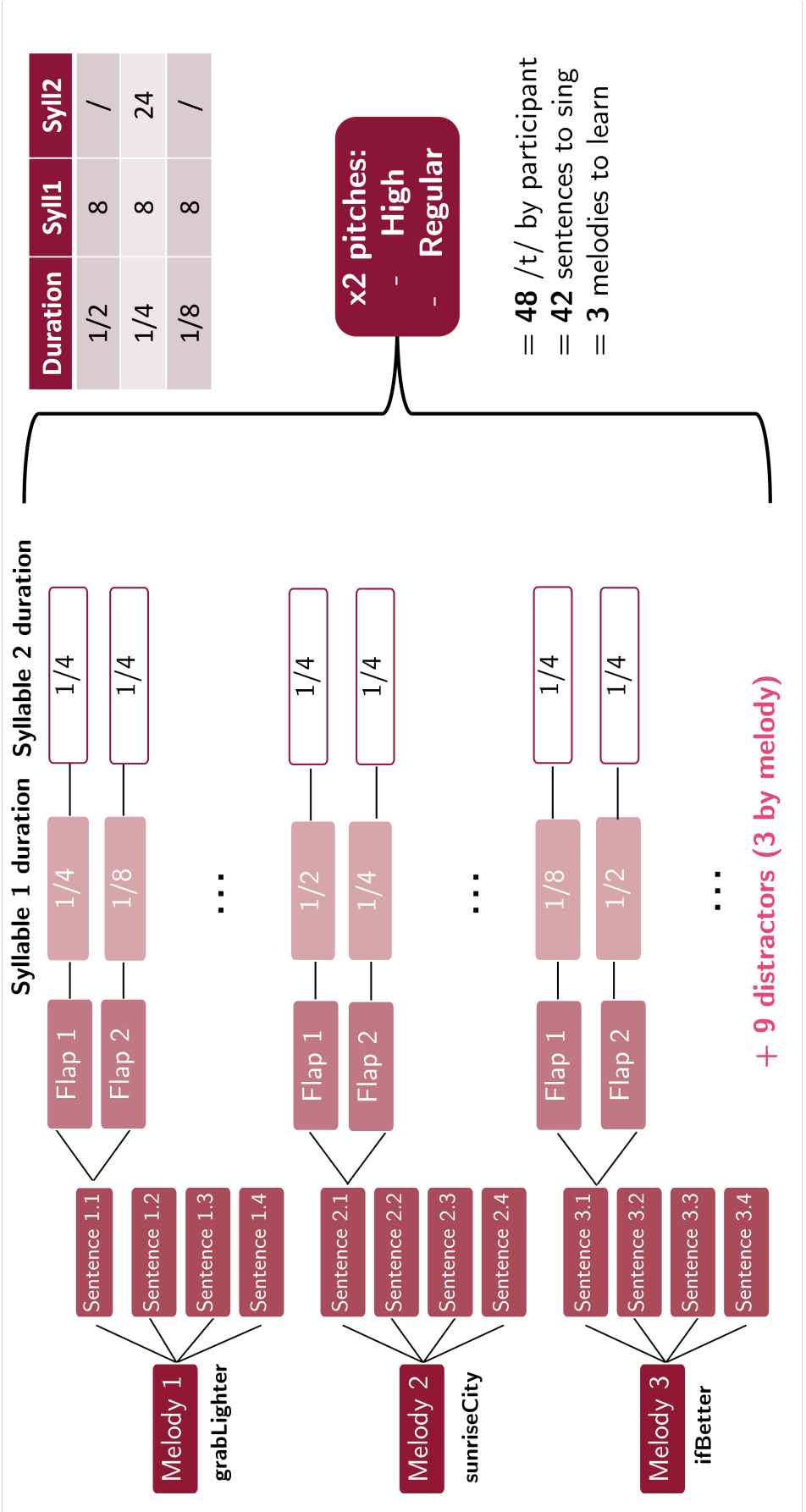


Figure 10.5: *Diagram of stimuli conditions for Production Experiment 2.*

Display

The display was identical to the one described in Section 9.2.3, Chapter 9. The script was created using Psychtoolbox-3.0.19 (Brainard, 1997) running on MATLAB R2024a (The MathWorks Inc., 2024), and can be found in Appendix B.8, and Figure 9.4 provides a diagram of the experimental flow.

10.2.3 Material and experimental setup

Pseudopalate creation

The process of creating a custom-made pseudopalate for each participant involved multiple steps. Rose Medical Solutions Ltd., which supplied the LinguaGraph EPG system, offers a palate-making service where the client can provide a palate mold, which can be taken by any dentist, and they return a custom pseudopalate created in their laboratory in England. This option guarantees a fully-working, clinical-grade palate, but is also lengthy and expensive–£580 (about 669€), with a delay of one to two months. I chose instead the option of buying DIY palate-making kits, for about £120 (about 140€) each. These kits include 62 silver electrodes and their attached copper-wire leads, a connector board and its cover and heat shrink tubing. Within the available funding and timeframe, this option was the most practical. The first step was therefore finding a dental prosthetic technician who would be willing to learn how to create EPG pseudopalates. After contacting different providers, a partnership was formed with the Laboratoire Fraisse[4] in Paris which specializes in dentofacial orthopedics and the creation of prosthetics.

The experiment ran over the course of six months and participants were recruited on a rolling basis. Once a participant was approved for the study, we made an appointment to meet directly at Laboratoire Fraisse. There, an impression of the upper part of the mouth (upper palate and teeth) was taken (see Figure 10.6.a). For four participants, this was done by a lab technician using an intra-oral scanner, a wand-like device inserted into the mouth which creates a 3D image of an individual's dental anatomy in a few minutes. This procedure was minimally invasive and the entire appointment lasted overall around 30 minutes. A custom model was then 3D printed (see Figure 10.6.b) the technician could use as a base to create the acrylic pseudopalate and embed the electrodes according to the participant's own dental anatomy. There was an issue with the scanner for Participant s01, whose impression had to be done with alginate, a rubber-like material which molds to the mouth. This creates a negative impression of the dental anatomy which is then used to make a plaster cast. Though the impression

[4] https://laboratoire-fraisse.fr/

method is different, and arguably slightly more uncomfortable, the model that is obtained is, for all intents and purposes, similar to the 3D-printed one and does not affect the remaining steps of the pseudopalate-making process. The delay for the creation of each palate ranged from three weeks to one month.

Once the base for the pseudopalate was finished (see Figure 10.6.c), the next step was to wire the connections to the EPG palate, which involved soldering the leads of each electrode to the connector board following a specific correspondence guide (see Figure 10.6.d) provided in the DIY kits' instructions. This was done by the experimenter herself. To attach the leads, they were threaded through clamping holes on the connector plug. An electronic continuity tester was used to match individual wires to their respective electrode, and solder each wire to the appropriate contact on each side of the connector plug (see Figure 10.6.e) ensuring that the two points constituted a closed circuit that an electric current can flow through. The process was quite time-consuming since it had to be repeated for each of the 62 electrodes, and took around two hours per palate. Once complete, a plastic sleeve was placed over the connector plug and each end was sealed with hot glue, and the finalized palate (see Figure 10.6.e) was ready for use.

Setup

The experimental setup and material used are described in detail in Section 9.2.4. Additionally, a LinguaGraph EPG system (Rose Medical Solutions Ltd.) was used for recording linguopalatal contact, as described in Section 10.1.2.

10.2.4 Procedure

The procedure for this experiment is described in Section 9.2.5, though some details pertaining only to the EPG experiment are addressed here. Participants were immediately fitted with their pseudopalates when they arrived for their session at Université Paris Cité. This meant they had around 30 minutes to adapt to the device before starting the experiment. This is less than the 45 minutes recommended in McAuliffe et al. (2008), but time constraints did not allow for more adaptation time. Participant s02 initially had difficulty adapting to the palate past that time period, but rapidly got habituated after a few practice trials of the experiment. During that time, participants read and signed the consent form (see Appendix B.10), and the protocol was explained to them.

They were guided into the recording booth where the pseudopalate was plugged into the electropalatography system. Participants were asked to flatten their entire

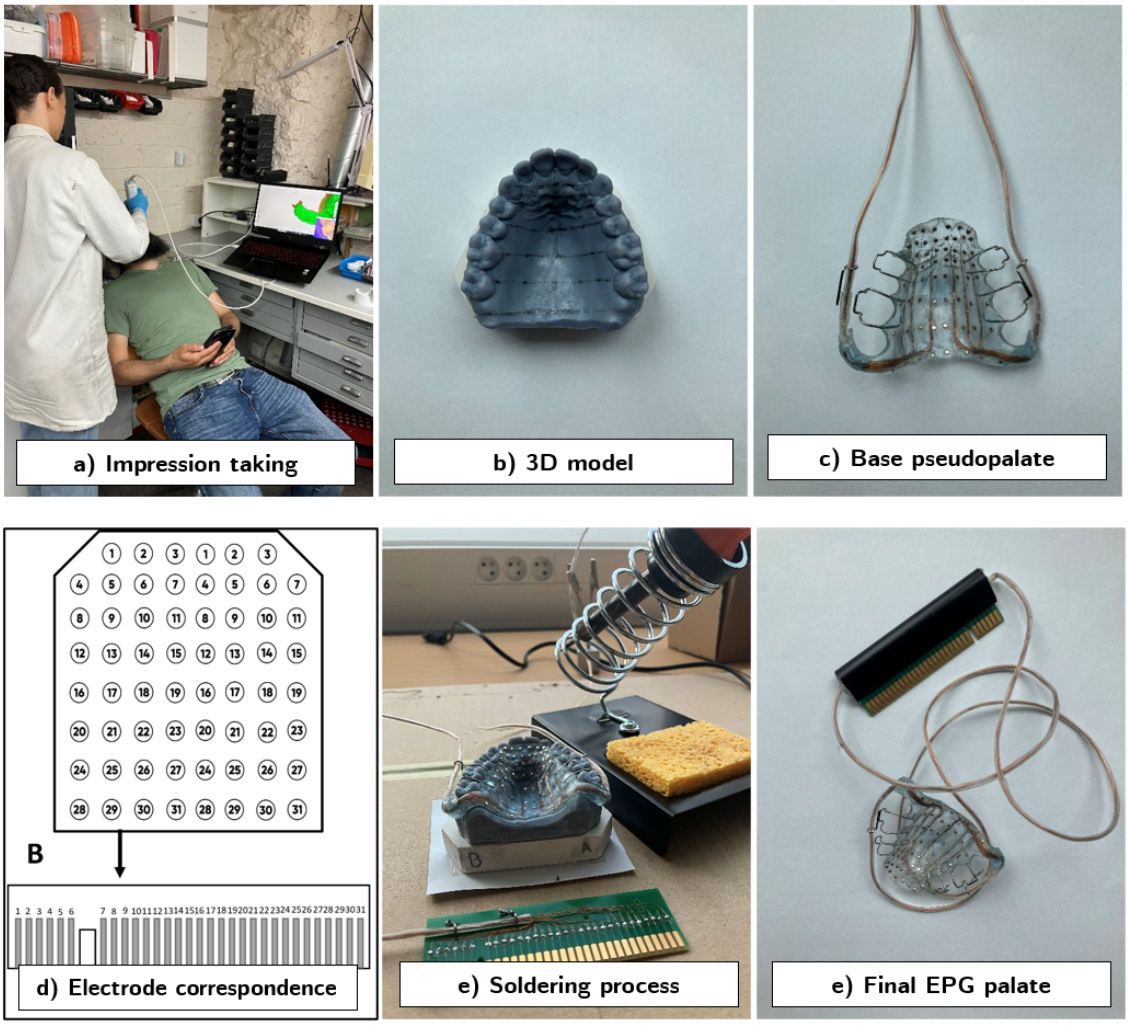


Figure 10.6: *Main steps in the EPG palate creation process.*

tongue against their palate to activate all electrodes as a way to adjust the sensitivity of the device using a dial located on the main unit. The sensitivity was turned up until all contacts were activated and not flickering. This was used to determine the baseline sensitivity for each participant, as the dial does not have numerical gradation to make sure it was the same level of sensitivity each time. Once set up, the experimenter returned to the control room, entered the specific reference notes for the participant, and launched the display script.

After the practice round, participants completed the first part of the experiment which involved a two-step process for each of the 42 trials: they had to **Listen** to the melody to get acquainted with the lyrics, then **Sing** it. This part lasted between 25 and 45 minutes depending on the participant, and the detailed experimental flow is illustrated in Figure 9.4. The second part of the experiment involved participants reading out loud in a normal voice the 21 sentences they previously sung.

They were then able to leave the recording booth and take off all the equipment. The third and final part of the experiment involved an informal semi-guided interview with the experimenter, with questions about the experiment (after which they were explained the aim), their singing style, and attitude toward accents in the singing voice. The full list of guiding questions can be found in Appendix C.1, though all of them were not necessarily asked. An abbreviated transcript of each conversation is provided in Appendix C.2. These will not be discussed explicitly in a dedicated section, rather, relevant passages will be mentioned throughout this chapter in order to shed light on specific trends. At the end of their session, participants were offered a 100€ compensation for their time.

All data was recorded using the icSpeech Professional edition software. The audio was recorded at a sampling rate of 44 100 Hz and a quantization depth of 16 bits. EPG data was recorded at 200 Hz.

10.3 Analysis

10.3.1 Preprocessing

Before analyzing the EPG data, a certain amount of preprocessing was necessary in order to isolate the specific tokens of interest. Unlike Production Experiment 1, the distractor sentences were not included in the analysis for this experiment. Similarly, where Experiment 1 looked at full sentences (spoken and sung), the majority of the present experiment focuses specifically on /t/ tokens, which is the reason for the discrepancy between the number of tokens in each. The audio and video from the icSpeech software

recording file (where audio, video and EPG data are all mixed in a .wmv format) were extracted in .wav and .mp4 formats respectively. Each audio file was imported in Praat (Boersma & Weenink, 2025), manually segmented at the target /t/ phoneme level, and annotated on multiple tiers:

1. **melody** – which of the three main melodies the token belongs to.
2. **sentence** – which of the twelve sets of lyrics the token belongs to.
3. **pitch** – high or regular pitch.
4. **syllable duration** – 1/8, 1/4, or 1/2.
5. **word** – which word does the token belong to.
6. **/t/ type** – whether the token is realized as a flap or as a stop (based on auditory analysis).

The resulting TextGrid served as reference for each participant for the subsequent analyses. The time intervals from the segmented audio were used to extract corresponding clips of the tokens from the video file, which could then be used for lip aperture analysis. Based on the stimuli presented in Section 9.2.3, the expected total number of /t/ tokens to analyze per participant was $n = 48$, including 8 tokens of 1/2 syllable duration, 8 tokens of 1/4 syllable duration and 8 tokens of 1/8 syllable duration, produced twice at two different pitches.

During the trials, no participant produced all 48 expected /t/ tokens, as several problems occurred during various stages of the experimental process, which impacted the final quality and amount of data available. Some of these issues were described in Section 9.3.1: since the experiment was self-paced (and no possibility of going back had been included in the experimental script), some participants skipped some trials; memory allocation issues occurred for sessions that lasted too long, which resulted in a loss of certain tokens. In addition, it came to light after his session that participant s00's EPG pseudopalate was faulty, to the point where the data was not salvageable even after post-hoc manipulation. Other issues inherent to electropalatography also occurred, but as they did not affect the number of tokens, they will be addressed in Section 10.3.2 which discusses EPG data analysis specifically. Table 10.3 provides an overview of the total number of tokens available for analysis for each participant for this experiment, s00 is in parentheses as his EPG pseudopalate was faulty. Though it is hard to determine the exact cause of this, his pseudopalate was the very first to be fabricated, and there was undoubtedly some element of trial and error in the process. It is likely there was a manufacturing defect, either at the electrode embedding or soldering stage of the creation, coupled with a too high sensitivity setting on the EPG unit. Post-experiment, it was clear that his linguopalatal contact data was unusually high, to the point where his whole palate was activated on average almost 80% of the

time throughout the entire session. The second highest whole-session average of all participants was s01 at 45%, which seemed more realistic. After an in-depth review of s00's contact patterns, it was determined that his pseudopalate was likely faulty and his electropalatography data therefore had to be discarded. He had been selected as a control participant in order to provide an American baseline to the data, so unfortunately no comparison was possible in the analysis.

Participant	Spoken	Expected spoken	Sung	Expected sung
s00	*24*	***24***	*46*	***48***
s01	24	**24**	47	**48**
s02	24	**24**	40	**48**
s03	24	**24**	47	**48**
s04	NA	**24**	36	**48**
Total (British)	**72**	**96**	**170**	**192**
Total (all)	**96**	**120**	**216**	**240**

Table 10.3: *Number of tokens available for analysis per participant for Experiment 2.*

10.3.2 EPG data extraction and reduction

As with the audio and video data, EPG contact data was recorded using icSpeech software. EPG raw data consists of a series of individual palate frames sampled at 200 fps. The resulting 64-bit EPG data was exported as a .csv file and further processed in MATLAB (The MathWorks Inc., 2024). The data for each frame was then converted back to a traditional-looking EPG matrix (see, for example, Figure 10.8) in MATLAB for further analysis using the EPG bitmap lookup table provided by the software.

There are multiple methods for data reduction in EPG studies (see Hardcastle, Gibbon, and Nicolaidis (1991)), one of the most popular being the Centre of Gravity (CoG) Index. CoG was created to indicate the location of the main concentration of activated electrodes across the palate, focusing only on the anterior-posterior dimension of articulation. However, this index is less relevant for a study that aims to compare the articulation of two alveolar consonants, which may not display any significant differences in terms of front-back contact. Furthermore, since lateral tongue bracing is also an important factor in the overall hypothesis, this analysis follows Fuchs and Perrier (2003) and Fuchs et al. (2006) in measuring different percentages of contact in various areas of the palate. Therefore, in addition to calculating the overall percentage of activated contacts across the whole palate, filters corresponding to specific anatomical areas of the palate were computed and added to the EPG Analysis script

as a further way of visualizing and analyzing the data. Fuchs and Perrier (2003) and Fuchs et al. (2006) describe anterior contact as the percentage of activation in the two frontmost rows of the palate and lateral contact as the percentage of activation in the two most peripheral columns on the left and right side of the palate. Following in their footsteps, an anterior filter was computed. However, it corresponds to the first four rows of electrodes on the palate as opposed to the first two only, as that is the area described as alveolar by the manufacturers of the LinguaGraph EPG system. Combining these two measures, a third measure capturing the "horseshoe configuration" of alveolar stops (Gibbon et al., 2007) was created, in order to measure both anterior and lateral contact–essentially functioning as an 'alveo-lateral-bracing' filter. It is referred to in this work as 'alveo-bracing' for convenience. This was meant to isolate the specific target contact pattern of interest, and thus avoid any activation in irrelevant parts of the palate which may be due do false contacts, over-salivation, or any other potentially unwanted information that may affect overall results. Using the timestamps from the segmented /t/ tokens in the TextGrid files (described in Section 10.3.1), corresponding EPG frames were isolated so they could be further analyzed.

Another measure analyzed was the variance of EPG contact, in order to examine both the levels of evenness of contact patterns, and the consistency of these levels across participants. Variance values for the average activation of each frame constituting a token were computed. For instance, if a /t/ occurred across 10 frames, then 10 variance values were obtained. They were then averaged to get a single global variance value per token. These values reflect spatial heterogeneity of linguopalatal contact during the token: higher values indicate more uneven or localized contact across electrodes (for example, a lot of concentration in one region and little or none in others), whereas lower values indicate more uniform or symmetrical contact across the palate.

Two methods of EPG data visualization were computed in MATLAB. One shows the EPG matrix as a heatmap, where the color gradient indicates the greater or lesser average activation percentage. The other is a function, `plotAnatomicalHeatmap`, inspired by the anatomical view option in icSpeech. The anatomical view is less precise than a heatmap in terms of average contact (there are only three different contact levels), but more realistic-looking, making the area of the palate touched with the tongue easily identifiable. Examples of both data visualization outputs are provided in Figure 10.7, for the sung /t/ token in the word *better* pronounced by Participant s01.

The full MATLAB script for EPG data extraction and reduction is provided in Appendix C.3, with its main functions defined at the end. However, before moving on to the analysis of the EPG data proper, one more issue concerning faulty electrode activation had to be addressed, which is done in the section below.

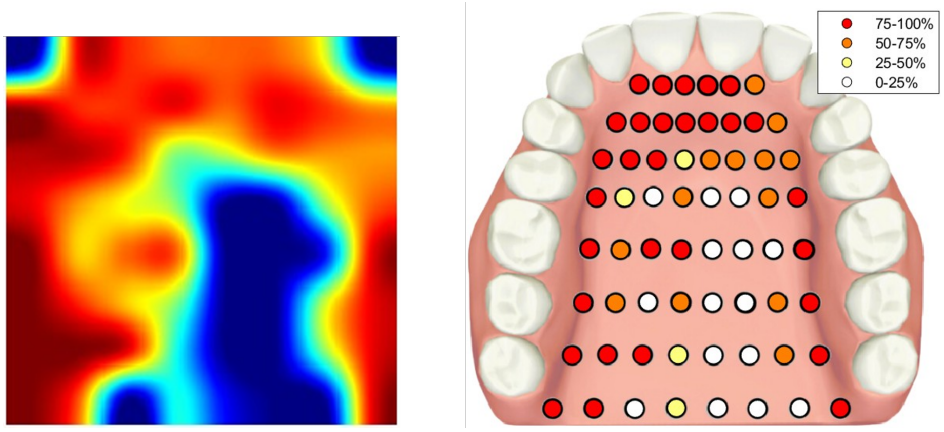


Figure 10.7: *Examples of EPG contact visualization (heatmap and anatomical) for intervocalic /t/ in the word* better *as sung by Participant s01.*

Electrode activation issues

Although finding a French dentofacial orthopedics laboratory for the creation of the pseudopalates and making the final adjustments myself best met the needs and constraints of this thesis, the novelty of the device, the difficulties of embedding electrodes, and the soldering process meant the end result was at times imperfect. In addition to participant s00's pseudopalate being faulty overall, the recurring main issue had to do with defective electrodes: some simply showed no activation at all, even when linguopalatal contact was present, whereas false contact seemingly happened with others who were activated despite not being in contact with the tongue. The first issue was easier to identify, as each electrode was tested with a continuity tester during the soldering phase (see Section 10.2.3), and a note could be made of the ones which showed no activation for future reference. The false contact issue was considerably harder to anticipate before the actual recording: as each was custom-made, pseudopalates could not be tested *in situ* beforehand, and despite participants going through an adaptation phase, excess salivation was likely the cause for a majority of these cases.

The icSpeech software has an option to reduce the number of false contacts in its integrated display, from which inspiration was drawn to resolve some of these issues. Data being already quite scarce, it was necessary to clean it up as much as possible, being mindful to avoid interference. Therefore, two MATLAB functions were created and added to the main EPG data processing script (which can be found in Appendix C.3, with both functions defined at the end of the main script). The function `fixContactlessElectrodes` identifies electrodes that are not activated but have at least 75% of activated contacts around them, and forces their activation. To the best of our knowledge, this type of manipulation has not been done with EPG data before, and there was therefore no precedent to determine how many surrounding electrodes

have to be activated to count the target one as defective and force its activation. After experimenting with different settings, it was determined that the threshold should be around 75% of electrodes activated around the target electrode in order to force its activation (see Figure 10.8.a). For an electrode not located at the edge of the palate, this corresponds to seven electrodes. This number was deliberately on the conservative side (the maximum amount of electrodes another one can be surrounded by being eight), so as to remove inconsistencies in the data without over-transformation. The second function, `removeFalseContacts`, deactivates a group of n or less electrodes that are isolated on the palate. Similarly to the first function, different settings were tested, and the threshold number was set to $n = 3$. This means that when three electrodes or less were activated on the palate, with no other contacts surrounding them, they would be considered false contacts and forcibly deactivated (see Figure 10.8.b). In both functions, the most conservative possible settings were used to avoid interference, though more extensive clean-up may have yielded clearer final results.

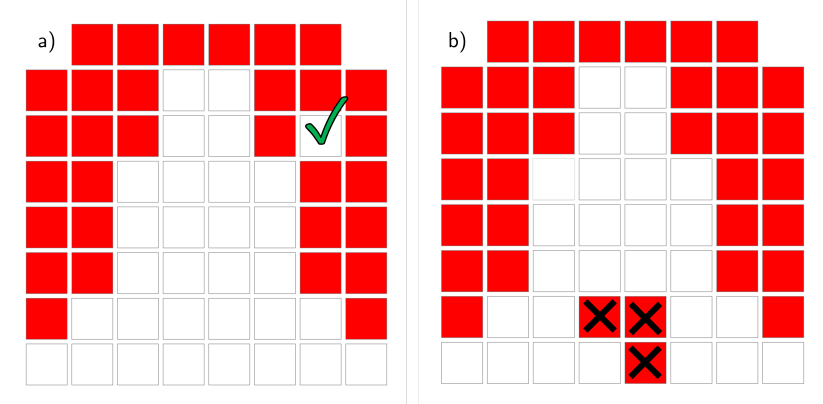


Figure 10.8: *Result of EPG data refining functions, with the function* `fixContactlessElectrodes` *applied in a) and the function* `removeFalseContacts` *applied in b).*

10.3.3 Pseudopalate normalization

Another issue that arose, directly related to the EPG palates, is that, despite expected individual variation–in the participants' own articulation but also salivation which may have caused false contacts–and the scripts described above to mitigate these effects, each pseudopalate seemed to behave very differently for each participant, to the point where it seemed the palate itself had to be considered a hard-to-control variable in its own right. Before the experiment, participants were asked to flatten their tongue against their palate to make full contact, and the sensitivity control was adjusted until all contacts showed constant activation on the live video feedback, without any

flickering on and off. However, the sensitivity setting is a dial without a precise scale, and therefore the setting was different for each participant. To make sure this was not an additional source of variability not attributable to articulation, all palates were tested by the experimenter afterward at the same sensitivity level. The level was adjusted using the experimenter's own pseudopalate. One by one, each palate was plugged into the EPG unit and a wet paper towel applied with equal pressure over the alveolar area was held in place for 10 seconds. The resulting contact averages for these calibration tests are displayed in Table 10.4. Though not perfect, the difference in resulting contact averages is relatively minor–it is therefore doubtful that further calibration is required to improve the data.

Participant	Mean alveolar contact (%)
s00	100
s01	92.95
s02	92.79
s03	99.6
s04	93
Experimenter	100

Table 10.4: *Mean alveolar activation from pseudopalate calibration tests.*

A relative activation ratio was also computed for each participant across the entire recording, in order to normalize within-participant pseudopalate activation data. All activation rates (whole palate, anterior and alveolar bracing measures) for each token were therefore divided by the overall rate of palate activation for the whole recording for each participant (indicated in Table 10.5). This resulted in a ratio where:

- 1 = equal to the participant's average contact in the session
- $>$ 1 = more contact than average
- $<$ 1 = less contact than average

This helped in accounting for how 'active' each palate was overall, and therefore to compensate for individual palate differences. For instance, s04's relative activation scores will be much lower than s01's, as his total activation percentage is very low. These measures are indicated by the (norm) precision after the variable names, as described in Table 10.6, which provides a summary of all variables used in the different analyses of this chapter, and their description.

Whole-recording activation (%)	s00	s01	s02	s03	s04
Whole Palate	78.12	45.55	20.04	18.91	8.31
Alveolar Bracing	86.07	53.44	28.95	21.08	10.4
Anterior	64.73	41.28	22.80	18.28	9.7

Table 10.5: *Whole-recording activation percentages by participant and region.*

10.3.4 Relabeling pitch

An issue that became evident during the experiment was that some participants did not seem to take the pitch condition into account when singing. It was made clear to participants that there were two pitch levels, which they were reminded of at several points in the process: during the familiarization phase with the melodies, in the practice round on the day of the experiment to confirm the two keys, and through the pitch cue provided before each melody. Participant s03 in particular only became aware after completing two-thirds of the experiment that he had not been paying attention to the pitch cues. It is possible there were too many other factors in the experimental design (including singing with a pseudopalate), so pitch was considered less important or forgotten by the participants, though no other participant reported a similar issue. Either way, to make sure the pitch condition would be reliable in the analysis, a post-experiment relabeling of the real produced pitch was conducted, to add to the final data on top of the pre-determined high versus normal conditions.

A Praat (Boersma & Weenink, 2025) script was created in order to extract the average pitch in Hz for each interval in the sentence tier of participants' TextGrids. The same settings were used as the for pitch extraction in Production Experiment 1, where the time step was of 1ms, and pitch floor and ceiling were set according to the participant's estimated vocal range (± 50 Hz). The resulting data was imported to R (R Core Team, 2024) and the `kmeans()` function from base R was used for k-means clustering, with $k = 2$ expected clusters based on the pre-determined high versus regular pitch categories.

The results obtained from this manipulation did not affect participants s01's and s02's data, who seemed to have succeeded in keeping with their high and regular pitch categories when singing. On the other hand, the data for s03 was considerably rearranged: where he was expected to have 24 and 23 tokens for high and regular pitch conditions respectively, his data was reclustered into 38 tokens for the regular condition (with an average of 163 Hz, or E3) and only 9 tokens for the high condition (with an average of 325 Hz, or E4). Similarly, s04's pitches were rearranged–where he previously had 22 and 14 tokens for high and regular pitch conditions respectively,

he ended up with 14 tokens for the high condition (126 Hz average or B2) and 22 tokens for the regular condition (225 Hz average or A3). That the count for each conditions are perfectly reversed is a coincidence, as the tokens themselves were not: some of the regular tokens remained regular and some of the high tokens remained high. Figure 10.9 shows the updated pitch clusters according to participant, where one point corresponds to one sung token. There are a few outliers, notably for participants s02 and s03, but they were not excluded due to the already-low total number of tokens.

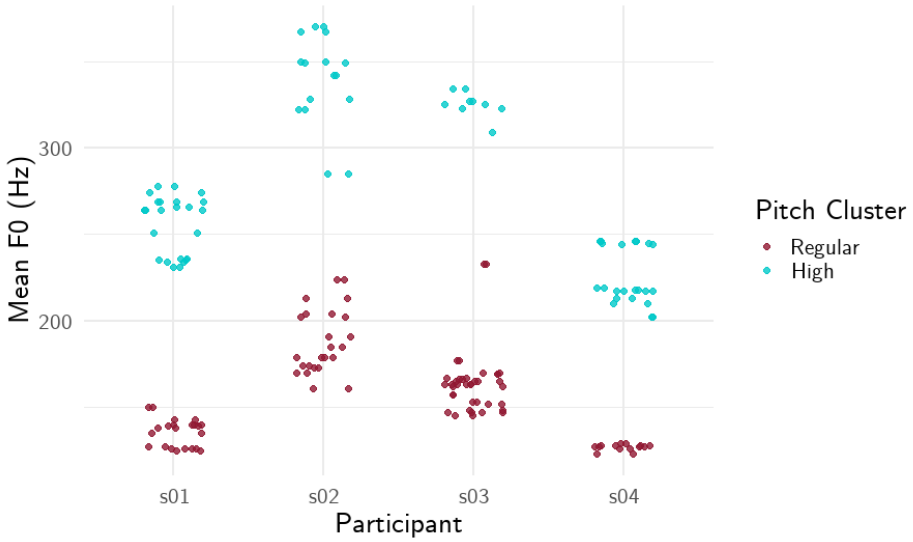


Figure 10.9: *Updated pitch conditions after k-means clustering by participant.*

10.3.5 Statistical analysis: a caveat

All statistical analysis in the following sections must be interpreted with caution. The small number of participants and analyzable tokens, along with the potential and hard-to-control effects induced by the pseudopalates make it difficult to find significant effects with statistical modeling. Though some statistical models are run, they should be considered as purely exploratory. Due to the nature of the data, a descriptive and individual approach will be taken throughout the Results and Discussion Section.

Statistical analysis was implemented in R (R Core Team, 2024) and RStudio (RStudio Team, 2025). The `lmer()` function of the *lme4* package was used (Bates et al., 2015) to fit linear mixed-effects models, and the `glm()` function was used for generalized linear models without random effects (used only in one instance). The p-values for fixed effects in the linear mixed-effects regressions were obtained using the *lmerTest* package (Kuznetsova et al., 2017), which applies Satterthwaite's approximation to estimate denominator degrees of freedom for t-tests. Table 10.6 summarizes all the variables

obtained from the different manipulations described in the current section, and that were used in the various models' syntax.

Variable	Description
/t/ Realization	Auditory perception of /t/ (flap or stop)
Pitch	The two pre-assigned pitch categories (high and regular)
Syllable Duration	The three levels of syllable duration
Relabeled Pitch	New pitch categories according to k-means clustering (high and regular)
Whole Palate Contact	Activation percentage for the whole palate
Whole Palate Contact (norm)	Relative activation obtained by dividing Whole Palate Contact with average activation across the whole recording
Alveolar Bracing Contact	Activation percentage for the alveolar and lateral areas of the palate
Alveolar Bracing Contact (norm)	Relative activation obtained through dividing Alveolar Bracing Contact by average activation across the whole recording
Anterior Contact	Activation percentage for the anterior (alveolar) area of the palate
Anterior Contact (norm)	Relative activation obtained by dividing Anterior Contact with average activation across the whole recording
Voice Condition	Speaking or singing

Table 10.6: *Summary of variables used in statistical modeling.*

10.4 Results and Discussion

Based on the hypotheses described in Section 10.1.4, the analysis involved several steps. Without looking at the EPG data and relying on auditory perception only, the first approach was to see whether or not flaps were elicited at all according to the different experiment conditions. Then, a general comparison of linguopalatal contact in singing versus speech was carried out, and in the last step, linguopalatal contact in the singing condition was analyzed. Results and discussion for each step will be given consecutively.

10.4.1 Eliciting flaps: results

All participants

The first level of analysis examined whether, based solely on the auditory identification of /t/ tokens as either flaps or stops, certain conditions elicited more flaps than others. Excluding s00, who consistently flapped as an American speaker, 39 of 170 word-internal intervocalic /t/ tokens (about 23%) were realized as flaps during singing. This was not equal in distribution across participants, however, as a majority ($n = 31$ or almost 80% of tokens) were flapped by the same participant, s02. The remaining 8 flapped tokens were mostly produced by s01 ($n = 6$), with participants s03 and s04 only flapping once. While the number of total sung flaps is too low to yield reliable results from statistical modeling, Figure 10.10 gives the proportion of stop and flap realizations of intervocalic /t/ according to the main conditions of the experiment–pitch and syllable duration.

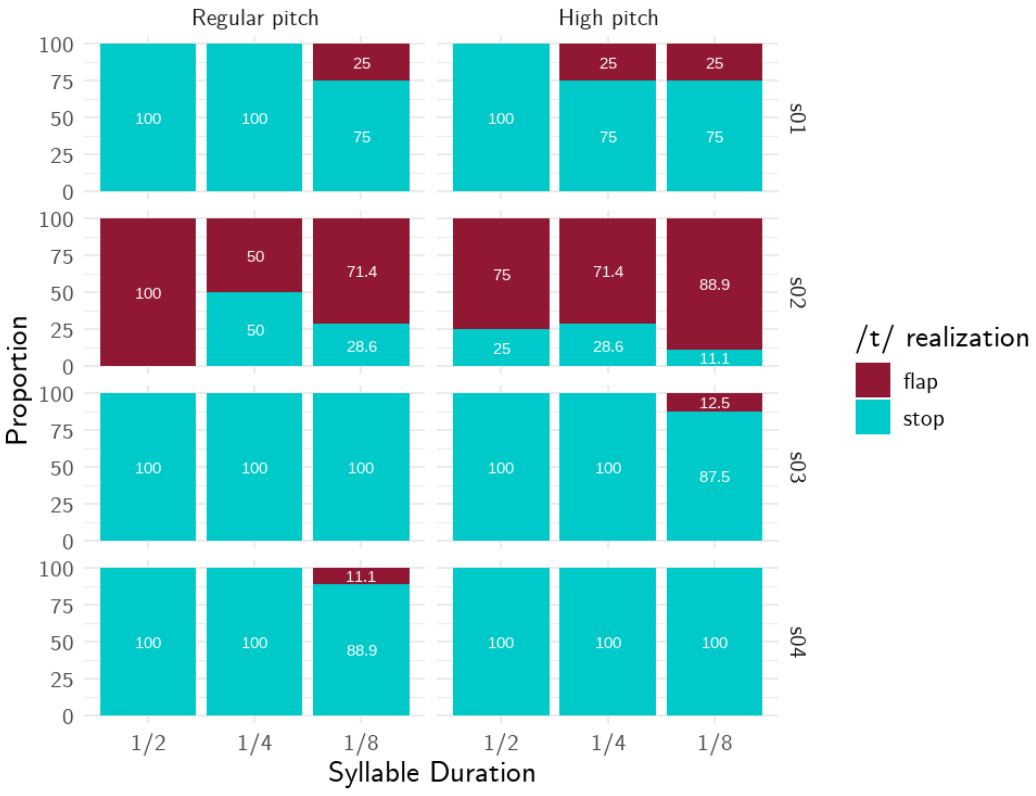


Figure 10.10: *Proportion of /t/ realizations in sung tokens according to pitch and syllable duration per British participant.*

In spite of marginal occurrences, all participants except for s02 show a clear preference for stop /t/ in their sung productions. Participant s02, on the other hand, displays a majority of flap /t/ realizations across all conditions, with a seemingly small bias to-

ward higher pitches. Looking more closely at the proportion of flaps and stops for each condition specifically combining all participants, Figure 10.11 does show a preference for flapping at higher pitches and shorter syllable duration (1/8 on the plot), though it is difficult to qualify this as a straight-forward trend at this number of participants and tokens.

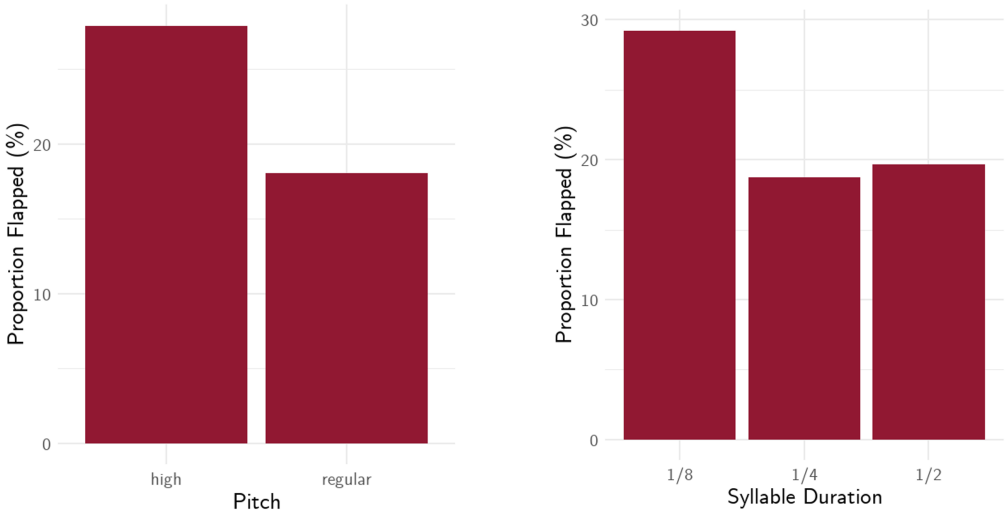


Figure 10.11: *Flapping proportion according to pitch and syllable duration.*

Flapped tokens

Although not forming part of the initial target of this experiment, examining which specific words were flapped yielded interesting patterns, which are discussed in more detail in the Discussion section. Table 10.7 summarizes the five most frequently flapped words and the participants who produced them. The remainder of the words not included in the table were either all flapped by s02 or not at all. The distribution of

Word	Times flapped	Who flapped
Better	4	s01 (both pitches); s02; s04
Party	4	s01 (both pitches); s02 (both pitches)
Beating	3	s02 (both pitches); s03
Matters	3	s01; s02
Bitter	2	s02

Table 10.7: *Top five most flapped words containing /t/ and who flapped them.*

who flapped these words aligns with the results above. Interestingly, only the word

better is flapped by three out of four participants. Other words in which /t/ was realized as a flap were only the case for two participants, even if they flapped across repetitions (hence why *party* is flapped a total of four times, but only by two singers), one of them consistently being s02.

A closer look at participant s02

Since participant s02 flapped the most out of all four British participants by far, a closer look at his specific data seemed warranted. Despite the small number of observations ($n = 40$) for a single participant, an exploratory model was fitted. This was not meant to draw generalizable conclusions but to illustrate possible trends in this participant's specific behavior. A simple generalized linear model was run on s02's subset of the data to examine the effect of syllable duration and pitch on the probability of flapping. The model included two predictors: syllable duration (with levels 1/8, 1/4, and 1/2) and pitch (high vs. regular). The outcome variable /t/ realization was binary, indicating whether a flap occurred. The model output is presented in Table 10.8.

Predictor	Estimate	Std. Error	*p* value
(Intercept)	1.599	0.745	= 0.031 *
Syllable Duration 1/4	-0.931	0.901	0.301
Syllable Duration 1/2	0.2236	1.002	0.823
Pitch Regular	-0.288	0.786	0.713

/t/ Realization ~ Syllable Duration + Pitch

Table 10.8: *Output of a generalized linear model predicting flapping as a function of syllable duration and pitch for participant s02.*

While the output of the model is not statistically significant, a few trends can be identified, illustrated in Figure 10.12, where error bars represent ±1 SE. For example, the negative estimates for syllable duration 1/4 and pitch regular may suggest reduced flap probability compared to syllable duration 1/8 and pitch high, in keeping with previous results.

Flapping in speech

Another important aspect of studying whether auditorily perceptible flaps were produced in singing was to verify any flaps occurred in the spoken data of British participants (excluding s00 once again, as he was American). As mentioned in Section 4.2, flapping is a phonetic feature primarily characteristic of American English pronunciation. It was therefore expected that no flaps would be produced in the spoken condition

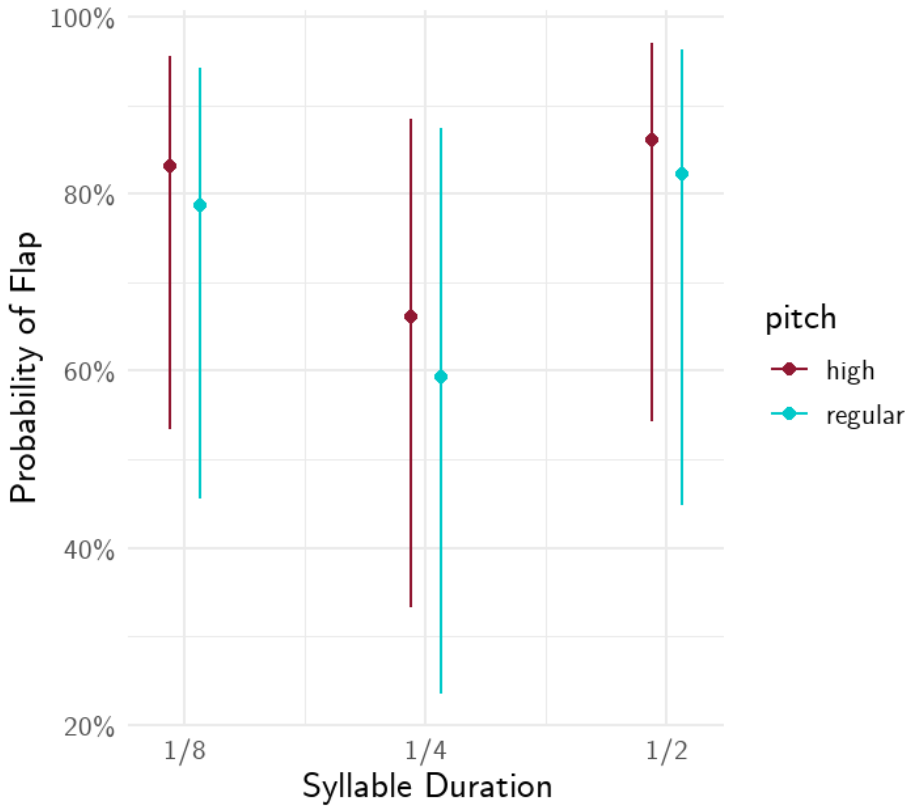


Figure 10.12: *Model-predicted flap probabilities by syllable duration and pitch for participant s02.*

of the experiment. However, all three participants whose spoken data was available (the one from s04 having been lost due to memory allocation issues), produced one flap out of the 21 spoken tokens available. What's even more striking is that all three flapped the /t/ token in the same word: *better*, which also happened to be the word most consistently flapped in the sung condition (see Table 10.7).

10.4.2 Eliciting flaps: discussion

Few auditorily perceptible flaps were ultimately produced in the experiment, and most of them are attributed to a single participant. It is therefore difficult to reach definitive conclusions because of this limited dataset, especially when it comes to statistical analysis. That being said, the above observations still bring to light certain interesting trends. For example, it does seem like syllable duration may have a subtle influence on flapping that a larger sample might reveal. Participant's s02's unique flapping tendencies compared to the other participants in the sung condition highlight the inherently individual strategies singers may resort to, and potentially even their specific attitudes toward their productions. Participant s02's specific profile may provide further explanation. An interesting element he volunteered in his post-experiment interview is that he was aware of singing with an American accent throughout his session, and even seemed to feel 'guilty' about it as he said "Because at one point I realized I was singing with an American accent and I'm British and I know I should be singing with a British accent ... I americanize myself." He provided an explanation for this, stating that, working full-time as a music producer, "there's a technical reason for that. I've been doing this for so many years just because I know that when you equalize, when you're equalizing the voice, the Ts are really aggressive. So I've learned to tone my Ts down ... I've learned to sing more Americanized to tone them down because it's more pleasant in the ear." The experimental context in a soundproof booth similar to a recording studio may have influenced his pronunciation by prompting habits linked to his career as a music producer. Interestingly, s01 who has also recorded several people (though not to s02's level), mentioned this issue as well: though he was not as certain of it as s02, he stated that it "might have come through." More broadly, this explanation is often mentioned in works on the style-shifting in the singing voice, whereby "[this] feature developed for a more microphone-friendly voice has become indexed as American through music recordings" (Jansen, 2022, p. 374).

Conversely, both participants s03 and s04 have stated in their own way their conscious choice to avoid certain features of the American accent, including flapping. Participant s03 openly admitted to trying to sound more British after listening back to his earlier productions and realizing that he "americanized a lot of the sounds." He

stated he "made an effort on the last album to actually bring [his English accent] back" because he "wanted to sound more natural, like [himself]." Though in his interview s04 initially stated he made a point to not pay attention to the way people pronounce things, he later remarked that he wanted to have "an individual voice which is not American." This does not necessarily suggest that he is trying to recover his original English identity through his accent, but rather he uses it as a device to set himself apart from other artists and create a character: "I probably want to sound like I sort of intellectually got a lot of baggage and I'm clever and all ... and so maybe that's the way of doing it: pronouncing it clearly and articulating and stuff like that. And let's face it, I'm not Maria Callas, that's not what it's about. It's just more of a character than sort of virtuosity, right." In a way, this attitude directly relates to Bell and Gibson (2011)'s description of the sociolinguistics of performance (see Section 2.2.1), where processes such as overshoot (exaggerating the characteristics of features of the target variety), may be helpful to interpret this specific case. In addition, it is also interesting to note that both singers had some underlying negative attitudes toward the American accent, despite explicitly stating otherwise: s04 associates it with being lazy, something that was also often repeated in the Perception Experiment survey responses, and s03 refers to it as "sounding dirty," which could also explain their much lower flapping rates.

Turning to the flapped tokens themselves, no specific pattern seems to appear that would explain why one word may be realized with a flapped /t/ over another, but the particular case of *better* is worth highlighting. Not only is it the most flapped word in sung productions (though this must be said with caution: it was realized as a flap in 50% of occurrences), it is also the one and only word that participants consistently flapped in the spoken condition, where no flapping was expected to occur. Additionally, the word *party* was also flapped four times (though only by half of the participants). An effect of lexical frequency may be speculated, and specific tokens will be examined in more depth in Section 10.4.7 below.

The speaking versus singing comparison also provided interesting, though unexpected, results. Indeed, if flapping occurs in speech (even if rarely) as was the case for the three participants with available spoken data, it can be assumed that it is therefore present in the individual's phonetic inventory and therefore may come out more easily in singing. This was not necessarily the case for the participants, as all of them flapped only once when speaking, but showed considerable individual variation in flapping rates when singing.

Lastly, this first exploration of the data relies on an auditory analysis of the tokens and their perception only. While there is a clear categorical distinction in a large majority of cases, some tokens remain more or less ambiguous. Most of participant s01's

realizations, for example, were classified as stops but were not necessarily as distinct or prototypical as other participants'. In any case, given the constraints of singing, the primary interest of Production Experiment 2 does not necessarily lie in how /t/ is perceived per se, but in what occurs at the articulatory level, particularly in terms of linguopalatal contact. Therefore if the first half of Hypothesis 8 and Hypothesis 9 in terms of elicitation of flap is not entirely confirmed, the actual amount of contact between the tongue and the palate remains to be studied. This will be addressed first as a comparison between speaking and singing, then looking at different factors in singing only.

10.4.3 Speaking vs. singing comparison: results

The second step of the analysis consisted in a comparison between speech and singing, much like the one carried out in Production Experiment 1, but this time studying linguopalatal contact with the help of the EPG system. Similarly to this previous speech versus singing comparison, participant s04 was excluded as his spoken data was lost during the recording. As indicated in Section 10.1.4, this part of the experiment involves testing the hypothesis that there is on average less linguopalatal contact in singing than in speech. Initially, average linguopalatal contact was studied across full sentence tokens, therefore including distractors (the number of tokens available for analysis in this case is the same as was presented in Chapter 9, in Table 9.3). To examine if voice condition (singing versus speaking) had an effect on linguopalatal contact across full sentences, a linear mixed-effects model was fitted in R, with whole palate contact (normalized) as dependent variable and voice condition as predictor. Random intercepts were included for participants and sentence. However, the model did not show any significant effect ($p = 0.335$), as can be seen in Table 10.9. A similar model was fitted previously using the raw whole palate contact percentages, which did not show any significant effect either ($p = 0.697$).

Predictor	Estimate	Std. Error	*t* value	*p* value
(Intercept)	0.761	0.048	15.596	< .001 ***
Voice Condition (sung)	0.061	0.064	0.0967	0.335

Whole Palate Contact (norm) ~ Voice Condition + (1 | Participant) + (1 | Sentence)

Table 10.9: *Output of a linear mixed-effects regression model predicting linguopalatal contact for the whole palate as a function of voice condition (speaking or singing).*

This may be due to the fact that the data is quite unbalanced (there are about twice as many sung sentences as spoken sentences for each participant). Figure 10.13

shows the average linguopalatal contact across the whole palate for the sung and spoken conditions per participant.

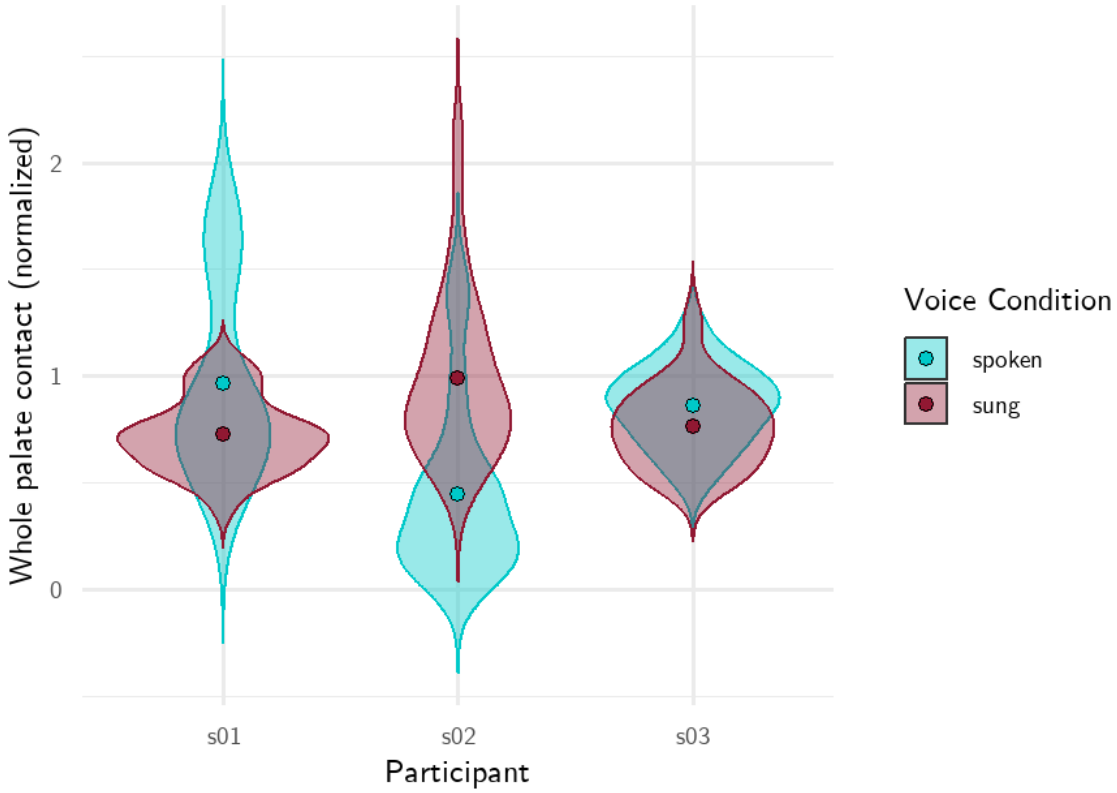


Figure 10.13: *Average whole palate contact according to voice condition (sung and spoken) per participant for full sentences.*

Participants s01 and s03 behave as predicted in Hypothesis 7, with on average less linguopalatal contact in singing than in speech, though the difference is slight when it comes to s03. Participant s02's distribution, on the other hand, reflects a reversed pattern with both the mean and most of the sung distribution exceeding those for spoken, an unexpected finding.

To further investigate the patterns of linguopalatal contact between speaking and singing, it was decided to look at activation patterns for /t/ tokens specifically, in addition to the full sentences. Two extra models were fitted to the data (summarized in Tables 10.10 and 10.11). In both models, the predictor was voice condition with random effects of participant and word. The first had average whole palate contact (normalized) as dependent variable and the second had average anterior contact (normalized) only. Both models were statistically significant–the whole palate one slightly more so than the anterior one, at $p < .001$ and $p < .05$ respectively. Statistical outcomes were similar for both models when using raw percentages for whole palate contact ($p < .001$) and anterior contact ($p < .05$).

Predictor	Estimate	Std. Error	*t* value	*p* value
(Intercept)	1.442	0.138	10.422	< .001 ***
Voice Condition (sung)	-0.420	0.087	-4.816	< .001 ***

Whole Palate Contact (norm) ~ Voice Condition + (1 | Participant) + (1 | Word)

Table 10.10: *Output of a linear mixed-effects regression model predicting linguopalatal contact for the whole palate as a function of voice condition (speaking or singing) for /t/ tokens specifically.*

Predictor	Estimate	Std. Error	*t* value	*p* value
(Intercept)	0.968	0.111	8.665	< .001 ***
Voice Condition (sung)	-0.202	0.101	-1.998	< .05 *

Anterior Contact (norm) ~ Voice Condition + (1 | Participant) + (1 | Word)

Table 10.11: *Output of a linear mixed-effects regression model predicting linguopalatal contact for the anterior region as a function of voice condition (speaking or singing) for /t/ tokens specifically.*

Looking at the detail of the distribution per participant for average anterior contact, illustrated in Figure 10.14, different individual trends can once again be identified. In this case, s02 appears the most in line with Hypothesis 7, the mean anterior contact for spoken /t/ tokens being much greater than that for sung /t/ tokens. This is also the case for s01, but there is considerable overlap between the two voice conditions. Participant s03 shows a reversed relationship than what was expected according to the hypotheses, with on average more anterior contact in singing than in speech.

10.4.4 Speaking vs. singing comparison: discussion

In this second part of the analysis, the expected outcome was that there would be less linguoapalatal contact when participants are singing than when they are reading the same set of sentences. The results did not show such a straightforward relationship, depending on which stimuli were examined: whole sentence or /t/ tokens, and which region of the palate was contacted. Because singing involves longer phonation time for vowels (see Chapter 3), and since previous results from Chapter 9 point to a more open vocal tract configuration in song than speech, less contact would be expected in all cases and for all participants. We found instead that this was not always the case. To try and explain the discrepancy seen in the data, especially with participants s02 and s03, a further analysis was carried out regarding speech rate, with the idea that if speech rate differed a lot between participants, it may be the cause for contact differences–

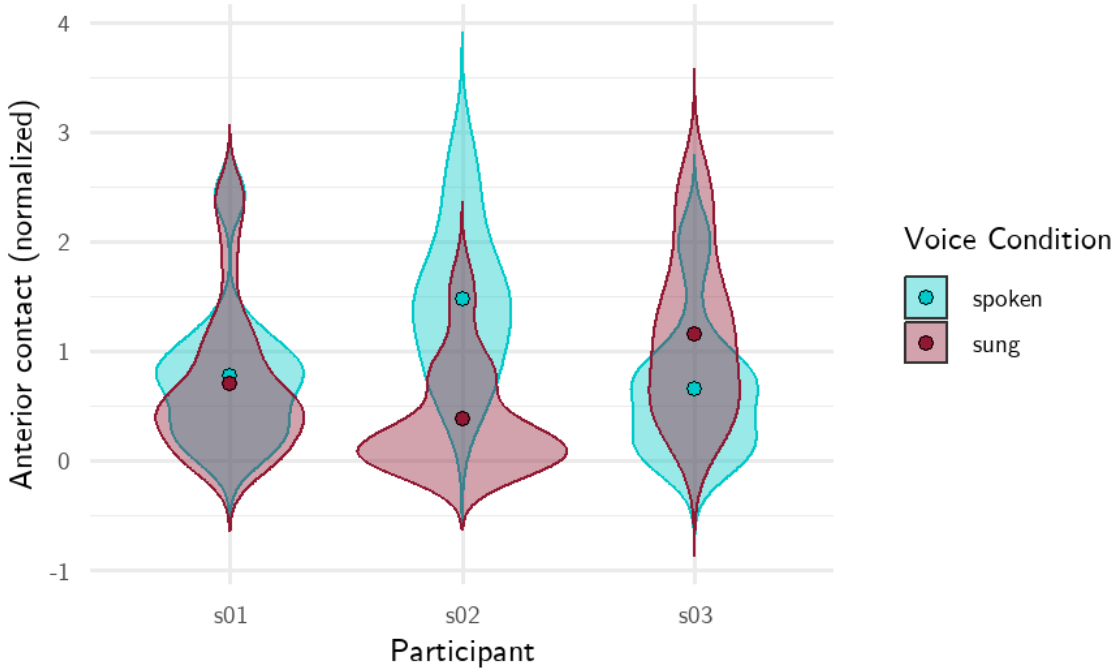


Figure 10.14: *Average anterior contact according to voice condition (sung and spoken) per participant for /t/ tokens specifically.*

for instance, faster speech rates in speaking may have induced decreased linguopalatal contact more so than singing by itself. Since the sentences were all fourteen syllables long, the total number of syllables they had to read was divided by the total time time it took for each participant to read them. Since participant s02 missed a sentence when reading, the speech rate was calculated over 20 sentences (280 syllables) for each participant. The same was done for all sung sentences (this time divided by each participants' specific token count). Since the melodies all had the same set tempo and meter (110 BPM, 4/4 meter), it was expected that participants would sing at similar syllable rates, but the measure was included for intra-participant analysis. Resulting measures are provided in Table 10.12.

Measures	s01	s02	s03
Spoken syllable rate (syll/s)	3.46	3.33	4.29
Average reading time per sentence (s)	4.05	4.18	3.26
Sung syllable rate (syll/s)	1.97	2.03	1.99
Average singing time per sentence (s)	7.10	6.9	7.03

Table 10.12: *Speech rate measures per voice condition for each participant.*

While participants s01 and s02 have fairly similar speech rates (around three syllables per second), participant s03 is notably faster, with a rate of 4.29 syllables per

second. This may therefore explain why s03 displays more anterior contact in singing than in speech, since his speech rate is considerably higher than both other participants, as well as almost twice the value of his own singing rate (almost two syllables per second). This difference may have been lost when averaging contact over all sentences (Figure 10.13), but was clearer when only looking at /t/ tokens (Figure 10.14). Furthermore, s03 stated in his interview that he felt like he wasn't actually singing, but just "saying words in tune" and that once he became aware of this, he started singing more. This may also have influenced his results and therefore could further explain the very close overlap in Figure 10.13 between his spoken and sung average contact for full sentences. Syllable rates are so similar between s01 and s02, so speech rate does not seem to explain why the latter displays more linguopalatal contact over the whole palate when singing full sentences as opposed to reading them, but less contact over the anterior region only when singing /t/ tokens than when reading them. This may be due to the fact that, as discussed in Section 10.4.1, s02 was by far the participant who produces the most flaps and, in keeping with the literature discussed in Chapter 4 and as shown in Figure 10.20 below, flapped /t/ is produced with less linguopalatal contact in the anterior region than its stop counterpart. In addition, s02 tends to show more consistent contact in the velar region (as can be seen in Figure 10.15). The combination of both factors averaged over full sentences as oppposed to /t/ tokens only may explain the discrepancy in speaking and singing contact. This is corroborated by the fact that, although he displays about 19% whole palate contact for all sentences when singing and 9% when speaking, he nevertheless presents only 14% anterior contact for all sentences in song as opposed to 17% in speech. Though this takes into account other alveolar consonants within the sentences, it is still helpful in putting s02's results into perspective.

10.4.5 Linguopalatal contact: results

Average electrode activation

As previously mentioned, the sparse nature of this data makes it difficult to obtain robust results from statistical modeling: the participants are too few, the variability in the equipment (despite normalization) too high, and there are only 170 sung tokens from British participants to study. Models were nevertheless fitted to get a first view of the data and test the significance of all possible combinations of the different variables that were computed (see Figure 10.6 for a description of each variable). This resulted in 12 different models, none of which, as expected, proved to be significant. Syntax and summary tables for each model can nevertheless be found in Appendix C.4.

Despite no statistical model providing significant effects, a descriptive analysis of the activation data can provide more insight. It must be noted that in all following figures and observations, the choice was made to use the most "reduced" parameters: unless otherwise stated, linguopalatal contact is shown for the anterior (alveolar) region only, and corresponds to its normalized version using the relative activation ratio described in Section 10.3.3. Since no particular statistical significance was found for any of the three filters described in Section 10.3.2, the anterior filter was selected for the sake of clarity, as it focuses on the area of occlusion. Pitch data corresponds to the relabeled categories as delineated in Section 10.3.4. In addition to discussing the various plots presented throughout the section, references will be made to Figure 10.15, which illustrates, for each participant, average electrode activation for all tokens corresponding to the different conditions of syllable duration and pitch. The diagrams were computed using the anatomical visualization script created in MATLAB (The MathWorks Inc., 2024), mentioned in 10.3.2–it displays four levels of activation, where a white electrode corresponds to an electrode that was activated between 0 and 25% of the time over all EPG frames recorded for the pronunciation of the whole token, yellow corresponds to 25 to 50% activation, orange 50 to 75% and red 75 to 100% average electrode activation rate. The palates are represented from above, but the view is flipped so that the surface that is touched by the tongue can be seen. As such, what appears on the left of the diagram actually corresponds to the right side of the participant. The levels of activation correspond to whole palate measurements, but an additional percentage is also provided in black for anterior contact only. The colors on the palate do not necessarily accurately reflect anterior activation only as the percentage provided is averaged only across the alveolar region (the first two lines of the palate indicated by a rectangle on the figure), as opposed to the entire palate.

Starting with the pitch condition, it appears that overall, there is less linguopalatal contact when participants are singing at higher pitches than at regular pitches. This can be seen in Figure 10.16, which shows the distribution of normalized anterior contact according to pitch. This observation is in keeping with Hypothesis 8, though there is significant overlap between the two boxes, suggesting the difference may not be extremely pronounced. Focusing on participants individually reveals that this difference varies greatly between subjects, and that the overall trend can hide dissimilar patterns among subjects. Figure 10.17 shows the distribution of normalized anterior linguopalatal contact across high and regular pitch conditions per participant. Whereas participants s02 and s04 seem to follow the expected pattern where higher pitches lead to less linguopalatal contact, participants s01 and s03 show a reversed relationship where high pitches seem to lead to more contact. That being said, the relative same

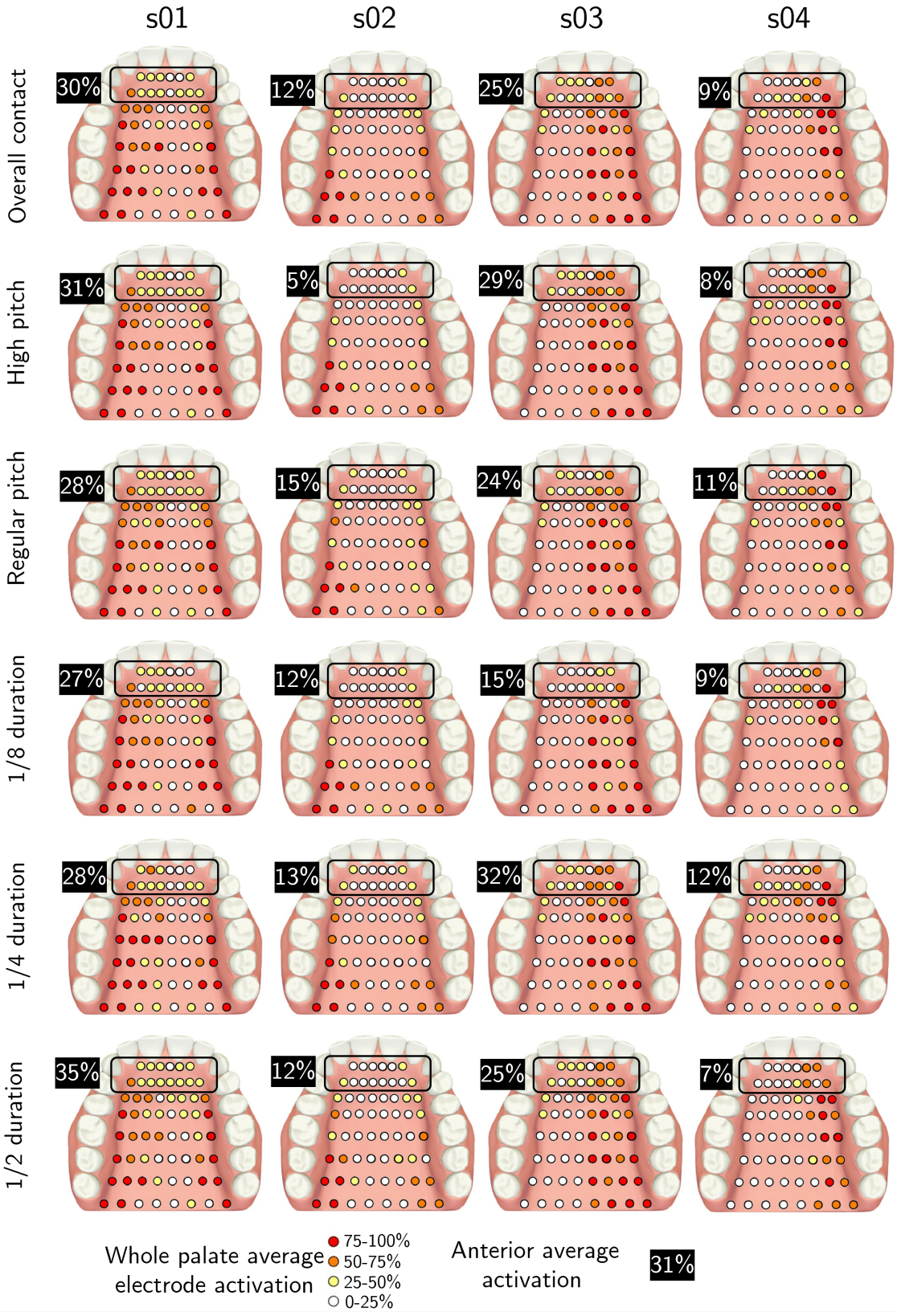


Figure 10.15: *Diagram of linguopalatal contact according to the different experimental conditions (pitch and syllable duration) for each participant. The black rectangle shows the alveolar area on each palate and the average activation rate of that specific area is also provided.*

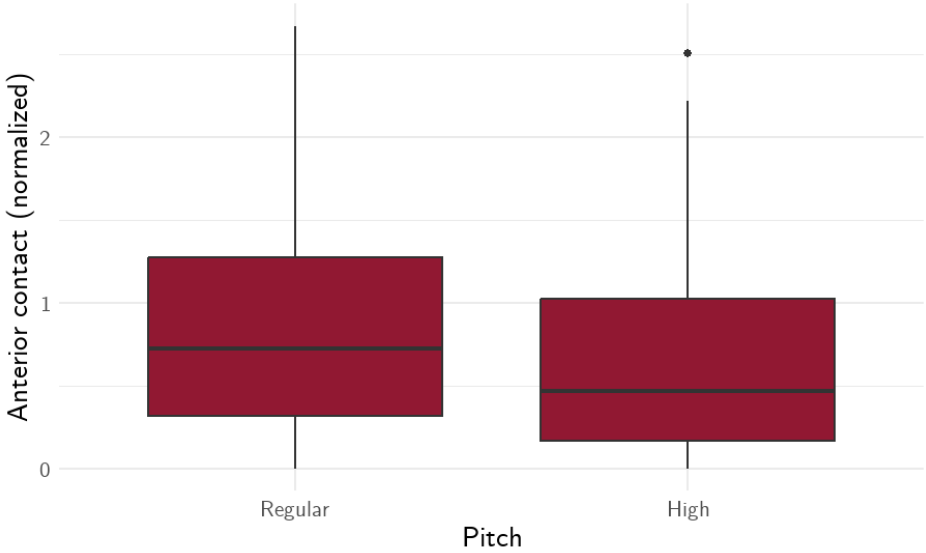


Figure 10.16: *Normalized anterior linguopalatal contact according to pitch.*

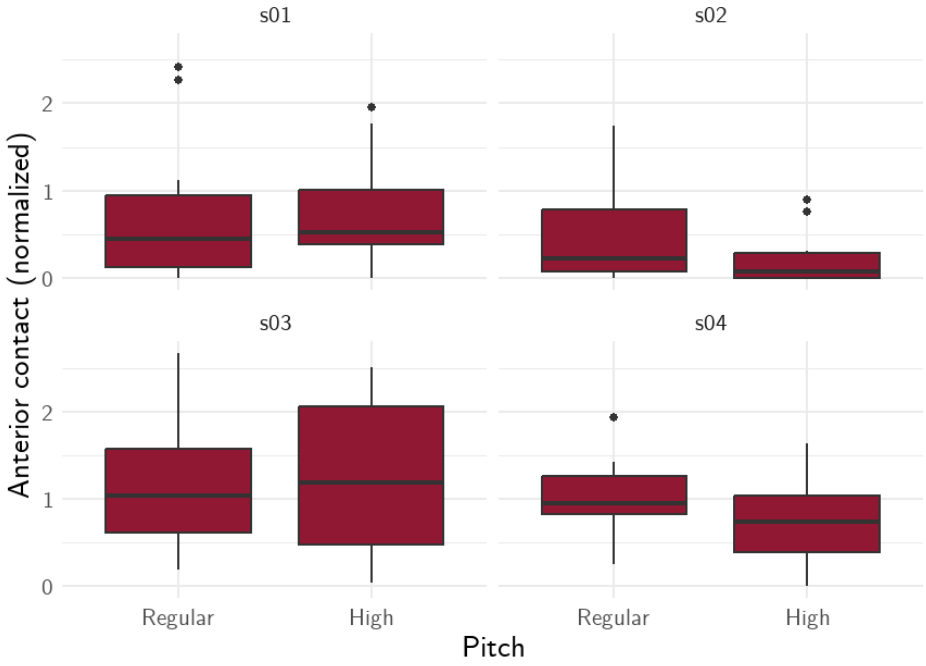


Figure 10.17: *Normalized anterior linguopalatal contact according to pitch per participant.*

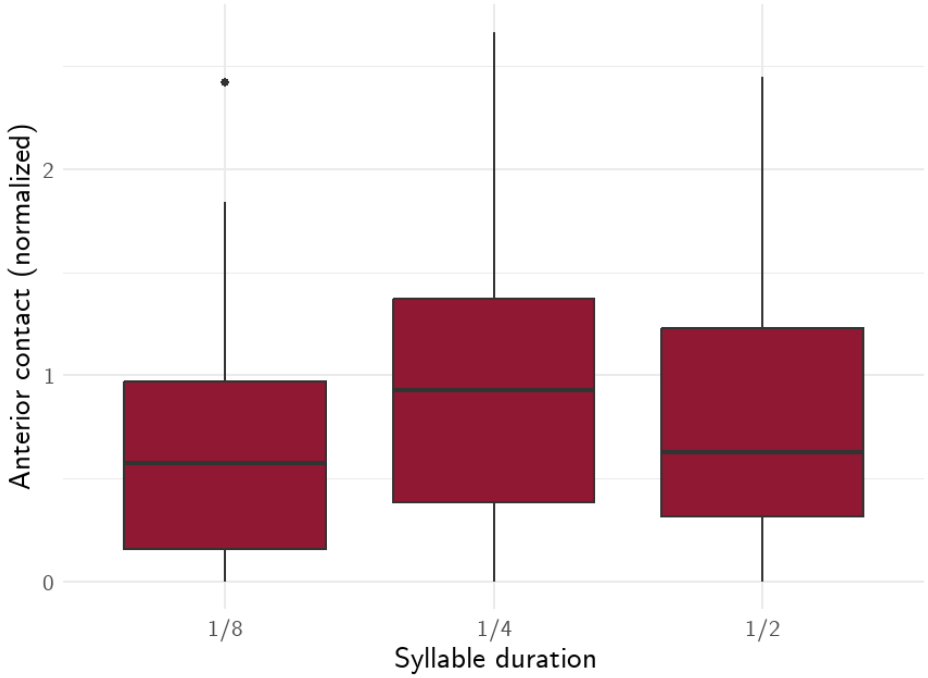


Figure 10.18: *Normalized anterior linguopalatal contact across three syllable duration (from shortest to longest).*

size of the distribution for s01 visually tends to suggest that there is in reality not much difference in contact at all between the two pitch conditions. This is corroborated in Figure 10.15, where anterior region contact only differs by 3%, and the activation patterns look very similar. This is also the case for s04, who shows the same difference in anterior activation rate. Similarly, s03's wide spread at high pitches indicates greater variability and less consistency in terms of /t/ contact patterns at higher pitches.

Turning to syllable duration, Figure 10.18 shows the overall normalized anterior contact across three syllable duration conditions (from shortest to longest). In keeping with Hypothesis 9, the shortest syllable duration (1/8) does seem to show the least amount of normalized anterior contact, and longer durations (1/4 and 1/2) appear to cause more anterior contact. However, the relationship is not linear as the middle duration 1/4 displays the highest contact, which actually decreases with the longest syllable duration 1/2. Much like for the pitch factor, a closer look at participant behavior reveals a variety of individual trends that the overall average hides. Figure 10.19 shows normalized anterior contact distribution according to the same three syllable duration categories, this time according to participant. Participant s03 and s04 actually appear to follow the general pattern shown in 10.18, where 1/4 syllable duration has the highest median contact of all conditions. Participants s01 and s02 on the other hand show fairly similar median contact levels across different syllable duration, suggesting that its effect is not consistent across participants, some being more influenced by duration

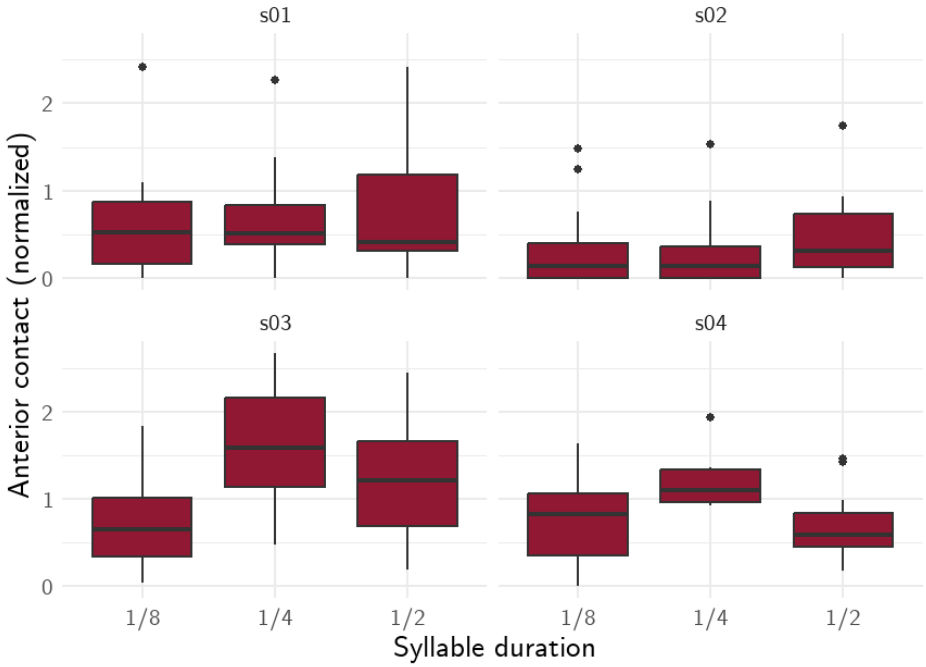


Figure 10.19: *Normalized anterior linguopalatal contact distribution according to syllable duration per participant.*

in their pronunciation patterns than others by this condition.

Lastly, since participant s02 was the participant who produced the most flaps in singing by far (31 out of 40 tokens), and because the American control participant s00's EPG palate was too defective to provide contact data for a prototypical flap, s02's data was further explored as a case study. His flapped realizations overall resulted in 16.3% of whole palate contact and 6.03% of anterior contact, as opposed to 20.6% whole contact and 17.9% anterior contact for stops. A Welch's two-sample t-test confirmed that stops exhibited significantly greater anterior palate contact than flaps for s02 ($t(10.92) = -2.73$, $p < .05$). Both anatomical figures and heatmaps for average whole palate electrode activation are shown in Figure 10.20, where the average difference in contact between stops on the left and flaps on the right can be seen. The hotter the color in the heatmap, the more average linguopalatal contact there is. For the anatomical map, the color code for the activation rates is the same as the one in Figure 10.15, where red means between 75 to 100% average activation and white corresponds to 0 to 25% of activation.

Spatial variance

Two levels of analysis were carried out on spatial variance. The first examined how evenly the tongue contacted the palate during the production of each token. In other

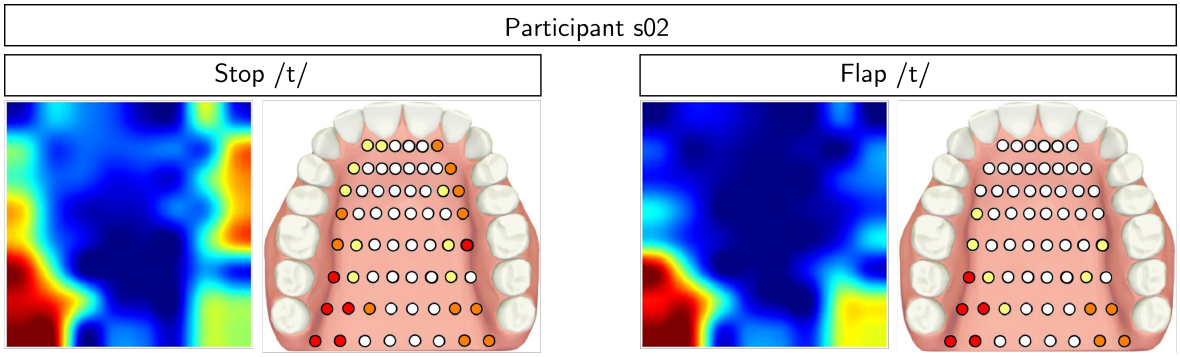


Figure 10.20: *Average whole palate activation heatmaps and anatomical maps for stops ($n = 9$) and flaps ($n = 31$) for participant s02.*

words, it looked at whether contact was distributed uniformly across the palate during the production of a single token. A Welch's ANOVA on per-token spatial variance values revealed a significant effect of participant ($p < .001$), confirming that contact differed from one participant to the next. Games–Howell post-hoc comparisons showed that all participant pairs differed significantly in mean variance ($p < .001$) except s02 and s03 ($p = .575$). The results are summarized in Table 10.13, where higher mean variance values correspond to more uneven contact–meaning contact is highly concentrated in certain regions. Participant s01 produced the most uneven tongue–palate contact patterns on average, whereas s04 produced the most uniform. This pattern is reflected in 10.21, where the heatmap for s01 shows large high-variance regions (in the hotter shades of red and yellow) compared to s04, whose heatmap is predominantly blue, and to s02 and s03, albeit to a lesser extent.

The second level of variance analysis focused on the consistency of linguopalatal variance. That is, whether they produced similar contact patterns every time or varied from token to token. Using the same per-token variance values as input, a Levene's test for homogeneity of variances revealed a statistically significant difference across participants ($F = 23.16$, $p < .001$), indicating that some participants were more consistent than others in their contact patterns. Standard deviation values of the variance are summarized in Table 10.13. These values indicate levels of consistency, where a higher standard deviation implies that participants are less consistent in their average variance. Participant s02 demonstrated the largest variability in his production of /t/ tokens, suggesting he was overall the least consistent in his contact patterns, whereas participant s04 proved to be the most consistent across tokens. Together, these two analyses show that participants systematically differed in both the degree and consistency of their contact patterns.

Heatmaps of average variance per participant can be seen in Figure 10.21, which show where the contact pattern varies the most between tokens for each participant.

Participant	Mean variance	SD of variance
s01	**1241**	424
s02	796	**572**
s03	664	325
s04	111	86.9

Table 10.13: *Mean spatial variance (average variance of electrode activation across all /t/ tokens) and standard deviation (standard deviation of per-token variance values) of spatial variance values per participant.*

Hotter areas (shades of red and yellow) correspond to electrode positions with high variation between tokens, whereas cooler colors (shades of blue and green) show electrode positions that have stable contact across tokens–this may mean either consistently high or consistently low contact. For the sake of comparison, average activation heatmaps for all /t/ tokens are included in the upper left corner of their corresponding spatial variance heatmaps. If a cooler spot in variance corresponds to a hot spot in average activation, this means the electrode is contacted a majority of the time–this is the case for the bottom left side of s01's velar area. If a cool spot in variance corresponds to a cool spot in average activation, then it is rarely contacted (see for example, the alveolar area of s02's heatmaps).

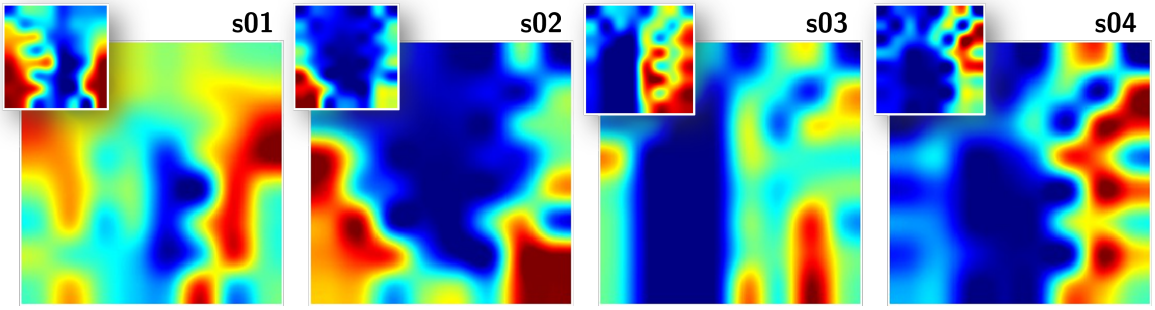


Figure 10.21: *Heatmaps of areas where the contact pattern is on average the most variable between tokens per participant. The hotter the color, the less stable the contact in that area across tokens. Average activation heatmaps are included in the top left corner for comparison.*

Dynamics of /t/

Aside from its ability to provide contact averages for specific pronunciation features, an advantage of using electropalatography is that dynamic trajectories of these features over time can also be examined at a high temporal resolution (200 Hz). The third step of the activation analysis focuses on the activation over time of the different

electrodes involved in the articulation of intervocalic /t/. According to the articulation of a prototypical stop /t/, as described in Section 4.1.1, there should be increased linguopalatal contact in the anterior/alveolar region–and likely over the whole palate as well, if lateral tongue bracing is considered–marking the constriction phase, followed by a rapid decrease of both marking the release. Since flapping is a lot more variable in terms of constriction, and may not even necessitate full closure (see the description in Section 4.1.1), the evolution of its linguopalatal contact is more uncertain. For all four participants, the trajectory in linguopalatal contact was computed across all EPG frames corresponding to each /t/ tokens. Resulting averages are shown in the line plot in Figure 10.22, which represents LOESS-smoothed average activation trajectories (for the whole palate and anterior regions) over normalized time, since raw tokens vary in duration. Participants s01, s03 and s04 all display a drop in palatal activation over time,

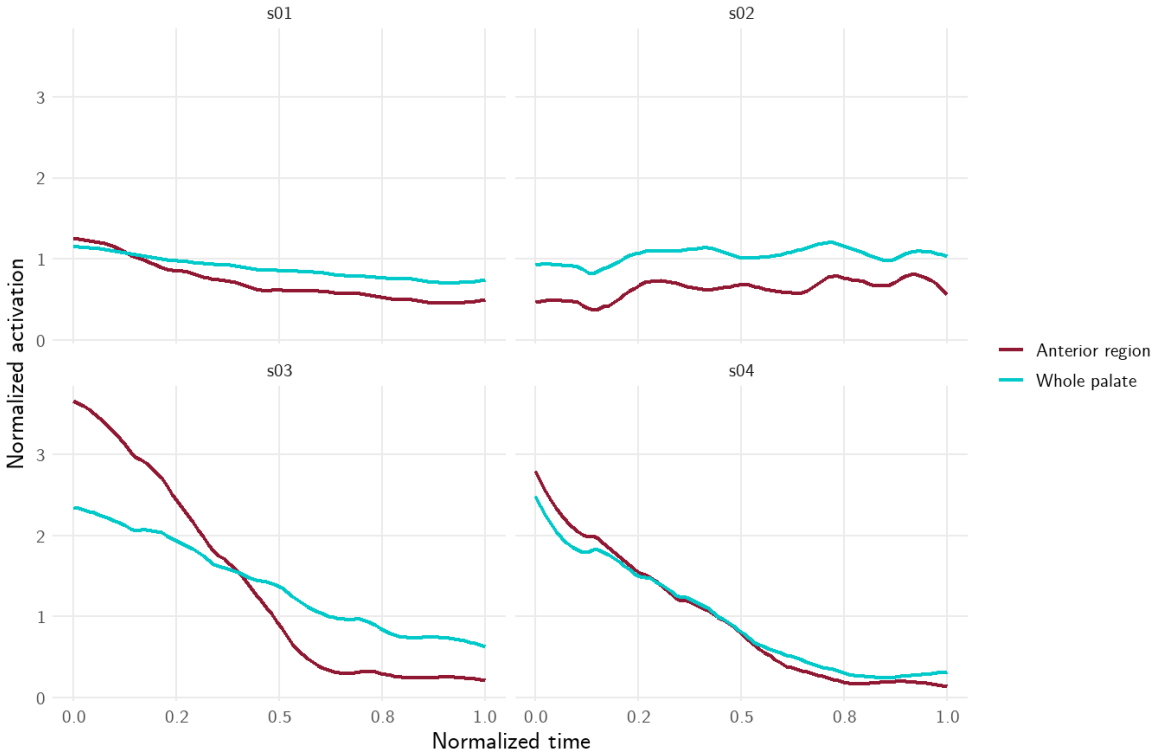


Figure 10.22: *Average linguopalatal contact trajectories over time per participant. Lines represent LOESS-smoothed values.*

consistent with the tongue making firm contact for /t/ closure and then releasing. This drop is particularly steep for s03 and s04, especially in the anterior region (dark red) suggesting rapid release form closure. The pattern is flatter for s01 which may suggest that contact is maintained a little longer before a more gradual release. Participant s02, on the other hand, shows a very different linguopalatal trajectory in his pronunciation of intervocalic /t/. His lines are stable over time, suggesting much smaller activation

changes and anterior region activation stays consistently below whole palate activation– being the only participant who flapped a majority of the time, this may reveal a specific flap trajectory, as will be considered in more detail in the discussion section below.

10.4.6 Linguopalatal contact: discussion

Average electrode activation

Before discussing the experimental conditions more closely, the overall results and a cursory look at Figure 10.15 already show a wide inter-participant variability. For anterior contact specifically, the minimum activation overall is 5%, (s02 - high pitch) and the maximum is 35% (s01 - 1/2 duration): the highest activation rate is about 7 times the lowest. s01 and s03 are on average bigger activators than s02 and s04, though the reasons differ for each. Whereas s02's contact patterns, where the anterior region is rarely activated, are not surprising since he actually produces flaps a majority of the time, s04 does show anterior contact. His lower rates are likely due to anatomical differences as opposed to /t/ pronunciation. This can also be linked back to the results of Production Experiment 1, described in Section 9.4, where he showed very low and uniform lip aperture measurements throughout. Regarding s03, and s04 to a lesser degree, contact patterns also show consistent asymmetry, skewed toward the right in the figure, and therefore his left. This tendency is very common: Verhoeven et al. (2019) found in a meta-study on EPG that lingual asymmetry in the articulation of speech sounds is substantial and occurs in about 83% of cases, 45% of them biased toward the left side.

Results for the pitch condition tend to verify Hypothesis 8 on average, where higher pitches lead to lesser linguopalatal contact, though per-participant results (shown in Figure 10.16) vary: both s02 and s04's results appear to confirm this relationship, as was previously suggested for s02 in Figure 10.12 showing model-predicted flapping probabilities of pitch). On the other hand, s01 maintains relatively uniform contact for both pitch conditions. Participant s03's greater variability of contact at higher pitches may reflect the increased difficulty of that pitch condition for him, which could have led to instability in his /t/ realizations. As mentioned in Section 10.3.4 on the pitch condition specifically, he also sung most of the stimuli at the same regular pitch despite cues, and only realized he wasn't singing at higher pitches halfway through. This resulted in only 9 high pitch tokens, which are likely not enough to provide a clear trend.

Syllable duration results on average also tended to show that a very short syllable duration before intervocalic /t/ leads to less linguopalatal contact, as posited in

Hypothesis 8. That being said, an interesting inter-participant trend is seen in both Figures 10.18 and 10.19, where the middle duration condition 1/4 induces the highest levels of anterior contact, which decreases again for the 1/2 condition. This may suggest that this medium syllable duration represents an optimal duration in prototypical stop /t/ articulation, whereas both shorter and longer durations may lead to altered articulatory patterns and less prototypical realizations. This is also corroborated by the fact that the only participant not displaying this trend is s02, who flapped most of the time regardless of syllable duration.

Overall, it appears that differences in activation within participants are quite small across all conditions, and their /t/ patterns remain consistent. The more remarkable differences appear across participants, who obviously have very different, idiosyncratic articulation strategies. Participant s02, who produced a majority of flaps in this experiment, tends to show the most salient differences.

Spatial variance

Turning to spatial variance analysis, results showed that participants differed both in how evenly they distributed tongue–palate contact for /t/ and in how consistently they maintained that pattern across tokens, revealing distinct articulatory strategies. Participant s01 displayed large regions of high variability, seen in his heatmap in Figure 10.21. His mean variance is high (1241) and standard deviation moderate (424). This means he tended to produce more uneven, or very precise, linguo-palatal contact patterns within each /t/ token, albeit in a relatively consistent way. Though he did realize some /t/ tokens as flaps when singing, it does not appear that they differed greatly from stops in terms of overall contact distribution. Participant s02 showed high variability in many areas, and his standard deviation value is the highest of all participants (572). His mean variance is relatively moderate (796), suggesting he was not as uneven in his contact as s01, pointing at different articulatory patterns across tokens. This may reflect less stable articulatory control or greater flexibility in achieving closure, consistent with the fact that he alternated between flaps and stops in his /t/ realizations. Both participants s03 and s04 proved to be more uniform and stable in their linguopalatal contact patterns. Specifically, s04 emerges as the participant with the most precise /t/ articulation out of all singers, with both the lowest mean variance (111) and standard deviation (86.9) of all singers. Notably, s03 and s04 both stated in their interviews that they are aware of their own pronunciation in singing and are careful and intentional with it, which may be reflected in their more uniform and stable /t/ articulations.

Dynamics of /t/

Despite individual variation in their /t/ token linguopalatal trajectories, participants s01, s03 and s04 all show a clear trend which corresponds to prototypical stop /t/ articulation with a sharp decline in anterior contact corresponding to the release, and a more gradual decline in whole palate contact. The larger gap between whole palate and anterior contact for s03 suggests a more fronted place of articulation for this participant, whereas s04 shows a more uniform constriction and release across the whole palate and the anterior area. Participant s02's contact trajectory is very different from the other three participants. His flat and overall stable trends for both anterior and whole palate regions suggest smaller variation in activation over time. More interestingly, the whole palate region (in cyan in Figure 10.22) remains more activated throughout the articulation of /t/, which is consistent with the absence of full alveolar closure that is typical with flapping. This is corroborated by lack of a sharp drop in activation, which would suggest a sudden release after defined closure, as is the case for the other three participants. It can be noted that s01 also produced some flaps during his session (albeit fewer than s02), which may also contribute to his flatter trajectory than s03 and s04, who only flapped once over 47 and 36 sung /t/ tokens respectively. Participant s01's more gradual release is also in keeping with a comment he made during the interview, where, like s02 though not to the same extent, he states that having recorded artists, he knows to tone down plosive consonant such as /t/ and /p/ as they "put massive spikes" on the waveform. Perhaps the gradual release seen in his average trajectory corresponds to his strategy for reducing the intensity of his consonants, without fully articulating a flap as opposed to a stop. On the whole, this seems to confirm that the presence or absence of a sharp drop in the trajectory of linguopalatal contact over time is a good visual marker of a stop or flap realization, which can be linked back to the fact that flaps tend to be realized with less tongue-palate contact overall.

10.4.7 A closer look at some tokens

Research constraints make a token-by-token study impractical, but observations about several individual tokens can nevertheless be made. Table 10.14 shows all tokens, ranked by their average anterior baseline value across all participants. Just like the token analysis in Section 10.4.1 on flap production, *better* comes out on top as both the word where intervocalic /t/ showed the least anterior contact (0.25 for the per-participant normalized value, or 5.92% for the raw value), as well as being realized as a flap the most (four times of out eight) and by the most participants (three participants out of four). As noted previously, it was also the only word to be flapped by all

Token	Average anterior contact (norm)	Times flapped
Better	0.25	4
Beating	0.39	3
Sweeter	0.40	2
Matters	0.51	3
Hurting	0.51	2
Later	0.54	1
Petals	0.55	2
Cheater	0.61	2
Shatters	0.66	1
Lighter	0.69	2
Photos	0.70	0
Fighting	0.72	1
Igniting	0.76	1
Created	0.80	1
Party	0.90	4
Started	0.91	2
Metal	0.94	1
Security	1.03	2
Bitter	1.06	2
Pity	1.07	1
Gutter	1.11	1
Letters	1.28	0
Vanity	1.32	0
City	1.34	1

Table 10.14: *Tokens according to anterior contact (normalized) and flap realization. Anterior contact data corresponds to the relative activation ratio described in Section 10.3.3. The tokens are ranked from lowest to highest average contact, across all participants.*

British participants in the read portion of the experiment, and also had the least overall average anterior contact for spoken tokens, equal to 0.319 for the normalized value or 9.83% in raw numbers. Since the lyrics containing the word (*if it's better for you, then I guess I'll see you later*) correspond to the third melody, the first syllable of *better* corresponds to 1/8 duration (see Table 9.2 for the lyrics and Figure 10.5 for the experimental condition). As the results have shown, this specific duration usually induces less anterior linguopalatal contact which may partly explain why *better* was flapped in singing. Other words matching this duration condition (namely *sweeter*, *photos*, and *fighting*) also tend to be produced with somewhat less linguopalatal contact, but not to the same extent as *better*. It also does not explain why participants realized it as a flap when reading. Though it is not the primary focus of this work, lexical frequency rate may provide an interesting perspective in the realization of this word. Indeed, out of all word tokens containing intervocalic /t/ in the stimuli, *better* is by far the most lexically frequent, according to both the Contemporary Corpus of American English (COCA) (Davies, 2008) and the British National Corpus (BNC) (Davies, 2004), with a frequency of 495 and 342 occurrences per million words respectively. To put things in perspective the next two words in Table 10.14 with the least amount of linguopalatal contact in singing are *beating* and *sweeter*, with frequencies of around 21 and 2 occurrences per million words for the COCA respectively and 17 and 1 for the BNC. This is in keeping with the idea developed by Bybee (2002) that reductive changes such as intervocalic /t/ being flapped tend to develop earlier and to a greater extent in words and phrases of high frequency. It also relates to the principle of least effort and economy of speech gestures as described in Chapter 3, where such sound change can be viewed as the "result of the automation of linguistic production" (Bybee, 2002, p. 268), whereby motor patterns such as speech become increasingly efficient as they are repeated.

Another interesting word worth examining is *party*. Although ranking fairly low in Table 10.14, it is the only other word to have been flapped a total of four times out of 8 potential times in the stimuli, with the caveat that it was only flapped by two participants twice, as opposed to three different singers in the case of *better*. Nevertheless, a few things are worth mentioning which may have influenced its realization. First of all is the fact that it was used as a verb in its sentence (*we're going to party*)–a typically American usage as opposed to British English which tends to use it as a noun only. This was not planned by the experimenter, but was actually pointed out by an English volunteer during the pilot session. Incidentally, right after the first occurrence of the word *party* during the experiment, s01 asked if he should be singing as though he were in a rock band, "with an American accent," suggesting a linguistic and cultural bias

associated to that token. It also relates back to the sociocultural perspective of specific genres influencing certain pronunciation patterns, as discussed in Part II.

Conversely, three words were consistently realized as stops in singing, including for participant s02: *photos*, *vanity* and *letters*. The last two show some of the highest values for normalized anterior contact (1.28 and 1.34 or 31.19% and 26.80% in raw values respectively), though *city* has the lowest overall value, despite the fact that it was realized as a flap once (unsurprisingly by participant s02). Both *letters* and *city* were words where the first syllable duration condition was 1/2, which may have influenced these results. The case of *vanity* is perhaps the most interesting: it was always realized as an auditory stop when singing, even by the American control participant s00. Participant s00 also realized one occurrence of *security* as a stop. Although his EPG pseudopalate activation data is unavailable for further analysis, this may point to a few things: both tokens have a quite high rate of average anterior contact overall for British participants (though *security* was actually flapped twice, by s02 both times). They both correspond to the same melody and the 1/4 syllable duration condition, which relates back to the discussion on electrode activation above (see Section 10.4.6) that this 'average' duration (as the mid point between 1/8 and 1/2) is actually the optimal duration for the articulation of a prototypical stop /t/. Furthermore, alongside *igniting*, which showed a medium to high value for anterior contact, they are the only polysyllabic tokens. Though they do obey the flapping rule, this may suggest the existence of a link that should be explored further.

10.5 Summary and Conclusion

The present chapter sought to explore how articulatory considerations inherent to singing may influence how singers realize intervocalic /t/ as a lingual consonant. Electropalatography was used as an experimental technique to measure linguopalatal contact in four different singers, participating in a karaoke-like recording session. The general hypothesis behind this analysis was that the more extreme or difficult the singing *context*, the more likely a flap or flap-like realization of word-internal intervocalic /t/ is to occur. Based on this assumption, this chapter examined the following points:

- **Speaking versus singing** - Due to physical and articulatory characteristics of the singing voice such as greater buccal aperture and vocal tract, singing would exhibit reduced linguopalatal contact overall compared to speech.

- **Syllable duration effect** - Shorter durations in the syllable preceding inter-

vocalic /t/ would increase the likelihood of flapping, and reduce anterior linguopalatal contact in the alveolar region.

- **Pitch effect** - Being linked to greater buccal aperture and increased articulatory effort, higher pitches would increase the likelihood of flapping, and reduce anterior linguopalatal contact in the alveolar region.

Across all analyses, it became apparent that while electropalatography is a valuable tool that provides rich articulatory detail, it presents notable challenges as a research method. The choice to import manufacturing of the palates to France and work with a local dentofacial lab made the process considerably easier in terms of time and expense. However, it also led to various technical issues caused by the novel nature of the technique. Despite implementing solutions and finding ways of working around these issues, some participant data was missed. Furthermore, the intrusive presence of the pseudopalate may have affected articulation to a certain extent, and heightened the awareness of being recorded in an experimental setting, which may in turn have impacted the naturalness of the performance. This may explain why the flapping rate found in this experimental setting was much lower than the one identified for both heavy metal and pop music (around 60% of /t/ realized as flaps) in the database analyses in Part II. These factors combined with the relatively small number of participants with individual singing styles limited attempts at statistical testing and generalization.

Nevertheless, in spite of these limitations, conclusions did emerge. Firstly, the most consistent pattern throughout all analyses was inter-participant variation. Differences in flapping realization and amount of linguopalatal contact were largely driven by individual articulatory strategies and stylistic choices. This notable inter-participant variation led to deeper analyses into variance and articulatory trajectories, confirming the deeply idiosyncratic character of sung intervocalic /t/ pronunciation. Each participant displayed their singularity in different ways.

s01 - Displayed moderate variability in contact patterns as well as relatively stable activation across conditions. Some flaps were observed in singing, and his /t/ trajectories were not always prototypical, though not entirely atypical. His anterior contact remained relatively unaffected by pitch and syllable duration, suggesting stable articulation.

s02 - The most prolific in terms of flapped realizations considered alone, s02 by himself confirmed the experiment hypotheses. His flaps were produced with lesser anterior contact than stops, variance and trajectory analysis showed a lack of stability in his realizations, reflecting a less precise and constricted articulatory gesture in keeping with previous descriptions of flapping (see Section 4.3.1).

s03 - Participant s03 exhibited the most prototypical stop /t/ characteristics out of all the participants, with a more fronted place of articulation and pronounced stability throughout. Unexpectedly, he had increased anterior contact in singing as opposed to speech, which may have been due to his considerably faster speech rate.

s04 - Though missing spoken data for comparison, the other analyses showed s04 as the most precise and consistent participant in his production of /t/, with stable and uniform realizations throughout.

More broadly speaking, certain patterns aligned with the initial hypotheses. For example, there was a slight tendency for shorter syllable durations (particularly the 1/8 condition) to correspond with reduced linguopalatal contact. There was also occasional evidence that higher pitches led to decreased anterior activation. However, these patterns were far from consistent across participants, and no statistically definitive conclusion could be drawn. Instead, the data reveal a complex relationship between broad phonetic and articulatory tendencies inherent to the singing voice and singers' own idiosyncratic production strategies. In addition to individuals' physiological differences, this experiment, and more specifically the interviews conducted with each singer at the end, also uncovered important links previously evoked in Part I and Part II of this work, between the way one sings, according to what genre or reference and how they want to be perceived. Both s01 and s02, the two singers who did flap, mention how their pronunciation is influenced by what they've heard: "Nine times out of ten, I'm mimicking stuff ... I would normally try and sing like the person I'm trying to imitate" (s01), or "As far as I'm concerned, I'm only singing like that because that's how I've heard other people. ... It's just that that's the way." This very clearly harks back to the ideas put forward by Gibson (2023) in his description of Pop Song English (PSE) as a standard way of singing that is not associated with any geographical or national connotations. Though he says he actively tries to do the opposite, s04 also makes this link explicit, when he states that "we've just been *abreuvé* [in French in the interview] with the American stuff, and the origin of the music we're playing is jazz and blues. And that's black American. And they've got a specific accent ... you can't escape as soon as you put something on like, a Chuck Berry song or something like that. You're there in the Delta or Chicago or somewhere." In addition, the two participants who were actively attempting to sound British both mentioned how it can be more difficult, "because it's not natural" (s04), despite being English due to the origins and influences of music, or because it requires "making a conscious effort" (s03), in keeping with Beal (2009), as discussed in Section 2.3.1. These findings suggest both the necessity of striking a balance between taking into account individuality in both

articulation strategies and attitude to what is being sung and conveyed, and scaling up the experimental scope to uncover potential trends.

11 Production Experiment 3: Magnetic Resonance Imaging (MRI)

This chapter describes the third and final production experiment, using structural real-time magnetic resonance imaging (MRI) as its principal method. In contrast to the two preceding experiments that focused on specific articulatory parameters using lip tracking and electropalatography, this study was designed as an exploratory investigation. Consequently, this chapter is primarily descriptive more than hypothesis-driven: it seeks to observe general patterns in vocal tract shape and jaw dynamics across singing and speaking conditions for two English popular music singers.

An overview of MRI and its application to articulatory phonetics is first presented, followed by a review of current research on the singing voice using this technique. Then, aims and predictions of the current study are outlined before detailing the methodology and data analysis procedures. The MRI images were used to analyze articulatory alignment and estimate jaw opening. Results are presented together with a discussion of their implications for understanding intervocalic /t/ production in singing and, more broadly, the differences between speaking and singing.

11.1 Background

11.1.1 Magnetic Resonance Imaging (MRI)

General description

Magnetic Resonance Imaging (MRI) is a non-invasive tomographic imaging method which records two-dimensional slices of a structure that can then be put together to reconstruct a volumetric three-dimensional view. MRI works using a strong magnetic field through which radio waves are emitted. This disturbs hydrogen atoms in the body that spin differently in different types of tissue, resulting in highly contrasted images of soft tissues, though it detects bone and teeth poorly. MRI was initially developed for medical applications, but has become an invaluable tool in articulatory phonetics research through its real-time variant, known as real-time MRI. The key distinction is that MRI captures continuous video rather than static images, enabling the study of dynamic movements.

In the last fifteen years, many studies have been carried out on physiological structures involved in speech productions thanks to MRI (D. Carey & McGettigan, 2017; Kochetov, 2020; Lim et al., 2021). At the vocal tract level, it allows the simultaneous visualization of the coordination of most articulators made of soft tissue during speech: the entire surface of the tongue, its underside, and other relevant structures such as the lips, the hard palate, the soft palate, and the pharynx. Nevertheless, there are a certain number of important limitations to the use of this imaging method. Firstly, it is not readily available: a partnership with a hospital or private MRI clinic must be established, and unless the machine is used exclusively for research purposes, medical exams have priority of use. It is also expensive, the average cost for a one-hour acquisition being around 500€. Depending on the scanner, the temporal resolution may also be limited, which can be problematic when trying to visualize very fast articulatory movements. Despite all of this, it remains the only technique able to provide an image of the entire vocal tract, making it an invaluable technique for future research in articulatory phonetics (Kochetov, 2020).

MRI in singing research

Singing studies have previously been carried out using MRI, though there are some potential issues due to the constraints imposed by the technique. One example is covered in Traser et al. (2013) who investigated the effect of the supine position participants must adopt in an MRI scanner. This effect consists of a higher larynx and slightly more protruded jaw compared to a sitting position. Traser et al. (2013) concluded

these limitations are not severe enough to undermine the conclusions of MRI research, and there is no impact on the articulatory components implicated in the production of intervocalic /t/. Various research questions have been addressed using MRI techniques. Echternach, Sundberg, et al. (2011) and Echternach et al. (2008) looked at the effects of vocal registers on the vocal tract of male opera singers, as well as vocal tract shapes of female soprano singers (Echternach et al., 2010). Sundberg (2009) and Bresch and Narayanan (2010) specifically addressed the question of formant or resonance tuning in sopranos, and the articulatory maneuvers of the lips, jaw and tongue that result from increasing pitch. In results comparable to those of Production Experiment 1 on lip tracking, Echternach et al. (2016) carried out an MRI study on articulatory differences in singing, finding variation in lip opening associated with changes in loudness and pitch. MRI studies on singing have focused on classically trained singers, though Echternach et al. (2014) have also examine music theater singing, usually considered a hybrid of CCM and western classical music (see Section 3.2.2). MRI has also been used to study production patterns and mechanisms in beatboxing (Dehais-Underdown et al., 2023), as well as yodeling (Echternach, Markl, & Richter, 2011).

11.1.2 Aims and predictions

The aim of this experiment is, above all, exploratory. Few studies have been conducted on singing voice pronunciation and articulation using MRI, and fewer still looking at specific allophonic differences such as the one between a stop and a flap /t/ across different voice conditions. Unlike the other production experiments discussed in this work, this chapter does not test any specific hypothesis related to factors such as pitch, intensity or syllable duration. Rather, it seeks to assess the feasibility of using MRI to capture articulatory data in singing, develop techniques to record the singing voice during an MRI scan, identify any noticeable differences between sung and spoken productions of intervocalic /t/, and determine whether instances of flapping can be elicited and analyzed under these conditions. In addition, specific focus is placed on jaw lowering, in this case expected to relate to findings from Chapter 9 on lip tracking, namely on the greater overall opening of the mouth and vocal tract in singing as compared to speech. Though it does not pertain to singing, Mooshammer et al. (2003) and Tabain (2012) have both found that for the production of alveolar stops, speakers seemed to adopt a high jaw position. Yet, if singing tends to require an open configuration of the vocal tract, it may be more difficult to produce prototypical stop /t/ productions. In light of these considerations, the following hypothesis will therefore be tested:

Hypothesis 10 Jaw lowering is on average greater in singing than in speech and

therefore has an effect on the realization of /t/.

11.2 Methodology

11.2.1 Participants

The same inclusion criteria as detailed in Chapter 10 applied to participants for this experiment. In other words, they had to be male, English, and professional or semiprofessional singers. Despite concerns that the debrief of Production Experiment 1 and Production Experiment 2 would influence non-naive participants, the decision was made to re-contact previous subjects to take part in this additional MRI study, due to the previous difficulties in finding eligible participants. Participant s04 responded favorably. A modified version of the flyer used for Production Experiment 1 and Production Experiment 2 (see Appendix B.5 and Appendix D.1) was posted on social media groups and forums for Anglophone musicians in Paris. This is how the second participant, s05, was recruited.

Participant s04 – Participant s04 is described in Section 9.2.2, though it bears mentioning here that he has had several implants and other dental work done, which can cause artifacts in an MRI scanner and blur or distort the output. Nevertheless, his resulting images do not appear strongly impacted by this.

Participant s05 – s05 is a 65+ English singer from London, currently living in Paris. He speaks English and French, and has had a career as lead and backing vocalist in two different bands. He describes his main style of singing as pop/rock. He describes his accent as "Cockney with some traces of [his] parents' northern accents."

11.2.2 Stimuli

Stimuli were reused from Production Experiment 1 and Production Experiment 2, as described in Section 9.2.3 and summarized in Table 9.2. However, conditions were simplified so that participants would not exceed one hour inside the MRI scanner. As such, distractor sentences not containing intervocalic /t/ were discarded, and participants were asked to sing each sentence only once, in their normal range. Melodies were transposed according to their reference note, which happened to be E3 for both participants. They were also asked to read each sentence in their regular speaking voice, and in their best impression of an American accent.

Unlike Production Experiment 1, participants were given all sets of lyrics, alongside karaoke video files corresponding to each, ten days before the experiment. This was done to maintain a certain coherence between the two subjects: s04 having already participated in Production Experiment 1 a year prior, he was already familiar with the melodies and lyrics, so it was preferable that s05 also had a chance to familiarize himself with the lyrics beforehand. Providing stimuli prior to the session was also necessary due to the setting of the experiment. The MRI scanner can be very loud during acquisitions (around 80 dB) and with participants having to wear earplugs to protect their hearing, it was better that they were already acquainted with the melodies to avoid any concerns about background noise and unfamiliarity of the setting.

11.2.3 Equipment

MRI data was acquired on a 3T MR scanner (Vantage Galan 3T XGO; Canon Medical Systems, Tochigi, Japan). Mid-sagittal images of the vocal tract were acquired using a single slice 2D fast gradient echo sequence with TR = 2.8 ms, TE = 1.2 ms, BW = 781.25 Hz, FOV = 24 × 24 cm, flip angle = 5 degrees. In-plane resolution was 2.5 × 2.5 mm and slice thickness was 10 mm. Images were acquired at a rate of 10 fps with a 16-channel array Atlas SPEEDER Head/Neck Coil combined with a Flex SPEEDER Coil placed over the neck and mouth. MRI data was reconstructed with Deep Learning Reconstruction to denoise the images (AiCE by Canon). The audio was captured with a FOMRI III+ microphone by Optoacoustics, made out of MRI-compatible optical material. Participants were fitted with HP PI US 03 headphones connected to a Starter f MKII+ Amplifier (MR confon GmbH). The setup with the equipment is illustrated in Figure 11.1.

11.2.4 Procedure

The MRI data collected for this thesis was obtained through the OPT-IRM (Optimisation des séquences IRM avancées) protocol[1] study which aims to optimize advanced MRI sequences for image quality (reduction of artifacts, signal-to-noise ratio, acquisition time, stability of quantitative measurements) on a recently acquired MRI unit dedicated to research at the GHU-Paris Psychiatrie & Neurosciences. The study also aims to assess the feasibility of MRI research protocols using pilot subjects. The work described in this experiment was achieved in close partnership with the GHU and the Institute of Psychiatry and Neuroscience of Paris (INSERM).

Participants came to the research MRI center at the Sainte-Anne location of the

[1] https://ichgcp.net/fr/clinical-trials-registry/NCT05618990. Last accessed: 14/09/2025.

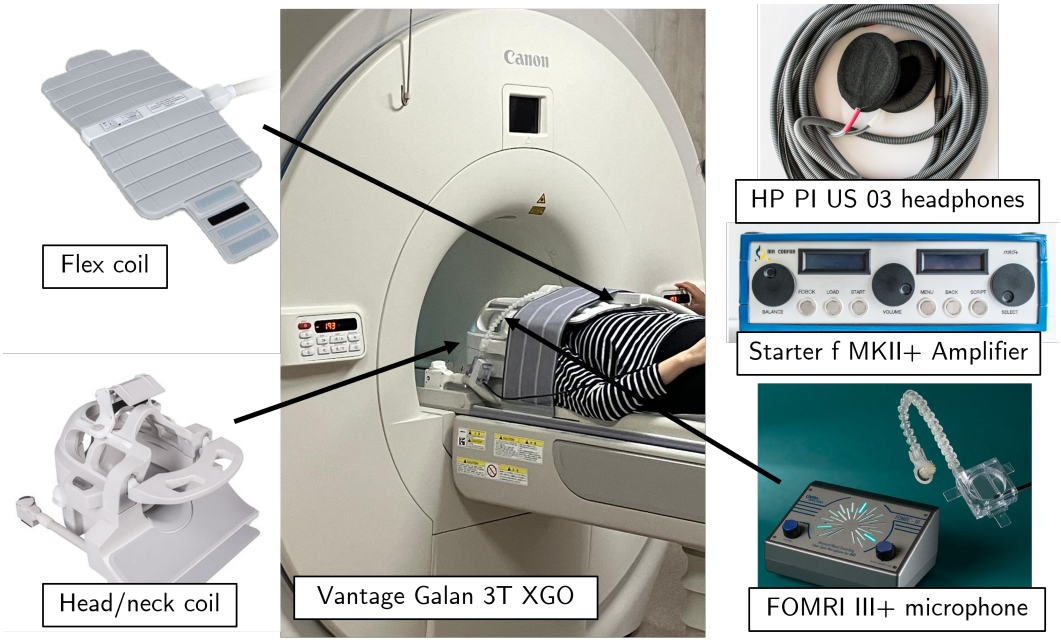


Figure 11.1: *MRI experimental setup.*

GHU-Paris Psychiatrie & Neurosciences. There, they met with the experimenter and the MRI technicians. Participants filled out a form to make sure they were MRI-compatible, and had an interview with a medical doctor for final approval to participate in the study. They also had to sign a consent form. Once officially included, they were positioned in the scanner in a supine position. The headphone was placed over the ears of the participant who was wearing earplugs, and set in place by extra padding. Participant s04 was also fitted with MRI-safe glasses corresponding to his prescription. The coils described in Section 11.2.3 were placed over the head of the participant, and the microphone aimed at the mouth of the participant was secured to the overall setup. Inside the scanner, a mirror placed at eye-level reflected a projection screen located inside the room. The participant could communicate at any time with the experimenter and technicians present in the control room, and was also given an emergency alarm button.

Instructions and stimuli were displayed on the screen in randomized fashion. Though similar to the experimental flow described in 9.4, the overall process was significantly simplified to cope with the constraints of the MRI. There were four steps involved for each twelve sets of lyrics, illustrated in Figure 11.2:

1. **Listen** - Similarly to Production Experiment 1 and Production Experiment 2, participants had to listen to the melody with the synthetic voice played over the karaoke video, as preparation for singing the melody and lyrics in the next step.

2. **Sing** - Similarly to prior experiments, the karaoke video from the first step was

repeated. The audio began with a tempo countdown and a pitch cue for the first note of the melody. Participants had to sing over the instrumental track, repeating the lyrics and melody. They could simultaneously hear a live feedback of their own voice through the headphones. They could not, however, listen to their own playback they sang.

3. **Read [BrE]** - Participants had to read the same lyrics they previously sang, with live vocal feedback.

4. **Read [AmE]** - Participants had to read the lyrics again in the best American accent they could manage, with live vocal feedback.

In total, the experiment lasted about two hours. The first hour covered signing the forms, getting the medical doctor's approval, giving instructions and setting up the participant in the scanner. The second hour covered the singing trials inside the MRI scanner. Participants were offered 150€ compensation.

11.3 Analysis

11.3.1 Data preprocessing

Audio was recording using the FOMRI III+ microphone, which has a built-in automatic noise reduction option, which was used when recording the audio files. However, since the MRI scanner is very loud, further post-hoc denoising treatment on the audio signal was necessary. This was done with the iZotope RX-10 software (Native Instruments), using the spectral de-noise tool. The files were then imported in Praat (Boersma & Weenink, 2025) and manually segmented and annotated on multiple tiers. The resulting TextGrids served as reference for the subsequent analyses:

1. **sentence** – the set of lyrics sung and/or spoken.
2. **word** – the word containing the token (×2 per sentence).
3. **/t/ type** – whether the token is realized as a flap or a stop (based on auditory analysis).

Concerning the images, MRI files are saved as DICOM files after acquisition. A MATLAB (The MathWorks Inc., 2024) script was created to convert these files into .avi videos. The script can be found in Appendix D.2. Once audio and video were separately processed, the files and their corresponding TextGrids were imported into ELAN (Max Planck Institute for Psycholinguistics, 2022) for further segmentation. When possible (this was not always the case due to the resolution of the MRI images), the exact frame where the occlusion of intervocalic /t/ occurred was identified through

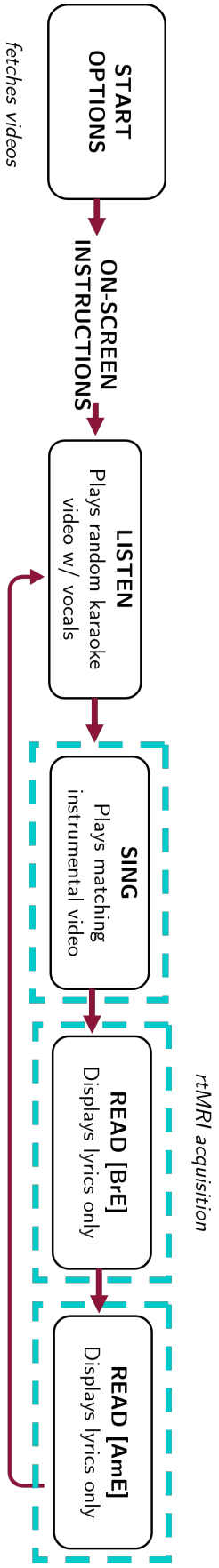


Figure 11.2: *Experimental flow for Production Experiment 3 (MRI).*

a frame-by-frame search using its segmented TextGrid interval. Frames were exported from ELAN, using the `Save current frame as image` option, at a resolution of 500 × 375 pixels. Tokens where no occlusion could be clearly identified were discarded, which resulted in a total of 24 lost tokens across all conditions. This was more frequently the case for the read stimuli, likely due to the increased speech rate that is too fast for the temporal resolution of the scanner. Additionally, some spoken audio files were lost during the experiment when the transition from the READ [BrE] to the READ [AmE] steps happened too quickly and the audio from the latter step was saved over the audio from the former step. This data loss occurred because resulting file names were saved as the name of the stimulus and the timestamp (down to the minute, but not including the second) of the recording, and certain transitions happened within the same minute. This occurred more often with participant s05 as I was more familiar with the process in this second run, and therefore launched each trial on the monitoring computer more quickly. Since the audio was not available for those tokens, the exact occurrence of /t/ could not be reliably identified solely on the basis of MRI videos. Seven videos had to be discarded from analysis for this reason. This resulted in a loss of 12 tokens for the READ [BrE] condition, 2 for s04 and 10 for s05. The detailed list of tokens available for analysis is provided in Table 11.1.

Participant	Sung	Spoken [BrE]	Spoken [AmE]	Total	Expected total
s04	22	19	18	59	**72**
s05	21	13	19	53	**72**

Table 11.1: *Number of tokens available for analysis per participant for Production Experiment 3.*

11.3.2 Articulatory alignment

As a first look into the resulting /t/ frames, a MATLAB script (see Appendix D.3) was created, in order to superimpose images for comparison across the different tasks. The first frames of the two acquisitions to be compared were aligned with the Registration Estimator app from the Image Processing and Computer Vision toolbox, using a rigid registration so as to preserve the positioning of stable anatomical landmarks, while compensating for small differences in head position from small movements between and during acquisitions. The resulting registration function between the first two frames was exported and applied to the two frames of interest, representing the moment of occlusion during the pronunciation of intervocalic /t/. The absolute difference between the two frames was computed in a single colored overlaid image. Areas

where both images perfectly align are gray, whereas differences are highlighted in green and magenta according to the different images. This color overlay is used to visually analyze articulatory differences between voice conditions or accents.

11.3.3 Distance estimation

This analysis follows the methods of Production Experiment 1, which examined lip aperture measurements in singing and speaking conditions. In this study, lip aperture was used as a proxy measurement of buccal aperture more generally. At the purely physiological level, buccal aperture is the result of a complex set of parameters including jaw mechanics (specifically the mandible) and the activity of certain muscles. Mandibular (lower jaw) gestures usually coordinate with tongue gestures in the articulation of stop consonants, and contribute to forming the constriction (Mooshammer et al., 2007; M. Stone & Hamlet, 1982). As the jaw actively contributes to the production of a variety of sounds both in singing and speech, MRI data can provide an interesting perspective into the mechanics at play in pronouncing intervocalic /t/ in singing. The following study provides an approximate estimation of jaw opening as another indirect measure of buccal opening. While classically trained singers learn to partially differentiate jaw lowering from buccal aperture (Sundberg, 1990), this was not the case for the two participants so jaw lowering can be used as an indicator of buccal aperture. Since there were issues finding the specific moment of occlusion for /t/, given the low temporal resolution of the MRI, vowels were privileged for this analysis as they last longer and would be more likely to be 'caught.' Using the same method as for /t/, the frame with the maximum mandibular opening for the vowel occurring right before the intervocalic /t/ in the tokens was extracted from the videos in ELAN (Max Planck Institute for Psycholinguistics, 2022). A MATLAB script was created (see Appendix D.4) to loop through all frames and open them one by one in the Image Viewer App for manual analysis. Two anatomical points were selected to estimate relative jaw opening: the very tip of the chin and the labial tubercle (the fleshy protuberance at the center of the upper lip). To maintain consistency, the points were selected by magnifying to the pixel level in the general region of interest. The pixel was then chosen based on its RGB[2] values, where the values had to be as close to 100 as possible, but remain above the 100 limit. This color corresponds to a low to medium intensity which approximately coincides with the edge of the face on the MRI frame. There is no exact demarcation between the background and the face, as the RGB values gradually lower

[2]Images were encoded as 8-bit grayscale stored as single-channel intensity values in the original DICOM files. The RGB representation in the exported frames simply replicates the grayscale values across channels.

to 0 which creates a blurred line effect. Though these points and selection process are arbitrary, the method was systematic across all frames, and ensured some coherence of the data. Then, the Euclidean distance between the coordinates of the two selected pixels was computed.

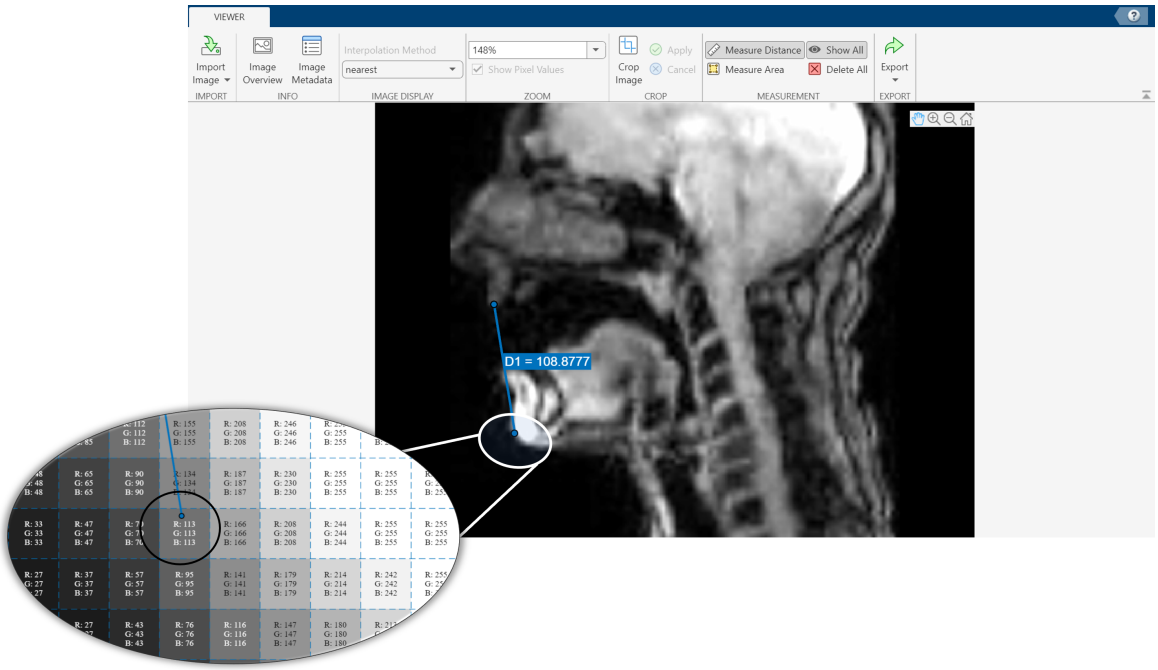


Figure 11.3: *Screenshot from the MATLAB Image Viewer App with magnified view of chin tip pixel and full view of the MRI frame with Euclidean distance indicated.*

11.4 Results and discussion

This experiment was mostly exploratory, and a part of the results presented here was drawn from careful visual inspection of the tokens under study. It is important to note the low 10 Hz temporal resolution of the MRI scanner used, as opposed to 200 Hz for the EPG used in Chapter 10, and up to 83 Hz in other studies involving MRI (Lim et al., 2021). It was therefore difficult to extract frames corresponding to the exact moment of /t/ occlusion. As stated in the methods, tokens were discarded when the frames skipped over the occlusion entirely, but some that occurred slightly before or after the moment of occlusion were kept.

11.4.1 Eliciting flaps

A preliminary auditory analysis focused on the perceived realization of /t/ tokens in the SING and READ [BrE] conditions. No flaps were produced by either participant in their spoken voice, but s04 flapped three times when singing *started*, *better* and *sweeter*. Participant s05 flapped once, singing the word *sweeter*.

Participant s04 produced more flaps in the MRI scanner than he did in Production Experiment 2 with the EPG, where only one /t/ token was realized as a flap. Incidentally, this was for the word *better*, which he flapped again during the MRI study. This may provide further evidence for the proposition put forward in Section 10.4.7 that since *better* has the highest lexical frequency of the tokens under study, it is more likely to be flapped (Bybee, 2002). That being said, this could also very well be a result of s04 having prior knowledge of the purpose of the research since he also participated in Production Experiment 2, though the sessions took place more than a year apart. For the word *sweeter* more specifically, the flap appears to result more from a moment of hesitation, which led to a mistake in the lyrics (he sang *such* instead of *much*), than from being familiar with the experimental context. This is also what seems to have happened for participant s05's unique flap /t/, which also occurred in the word *sweeter*. The participant added the word 'so' in the sentence *she's much sweeter than you*, and therefore had to compensate for the extra syllable by singing more quickly, which may have resulted in a flapped realization. The other sentences with the same corresponding melody and rhythm were sung without any mistakes by either participant. Both *better* and *sweeter* correspond to the same melody where the syllable duration was manipulated to last 1/8 of a note, in keeping with the previously identified trends in both Production Experiment 1 and Production Experiment 2.

11.4.2 Speaking vs. singing comparison: results

Alignment

Figures 11.4 and 11.5 show tokens in the sung and spoken BrE conditions for each participant, obtained from the alignment method described in Section 11.3.2. Gray regions indicate stable overlap between the two conditions, whereas colored regions indicate areas where there are differences in articulatory alignment between the two voice conditions. Green corresponds to the frame for the spoken condition of each /t/ token and magenta the corresponding sung condition. Since the images were registered based on the first frame of each video, additional noise due to head movements should be negligible. Boxes around images indicate the few tokens that were realized as flaps in the singing condition.

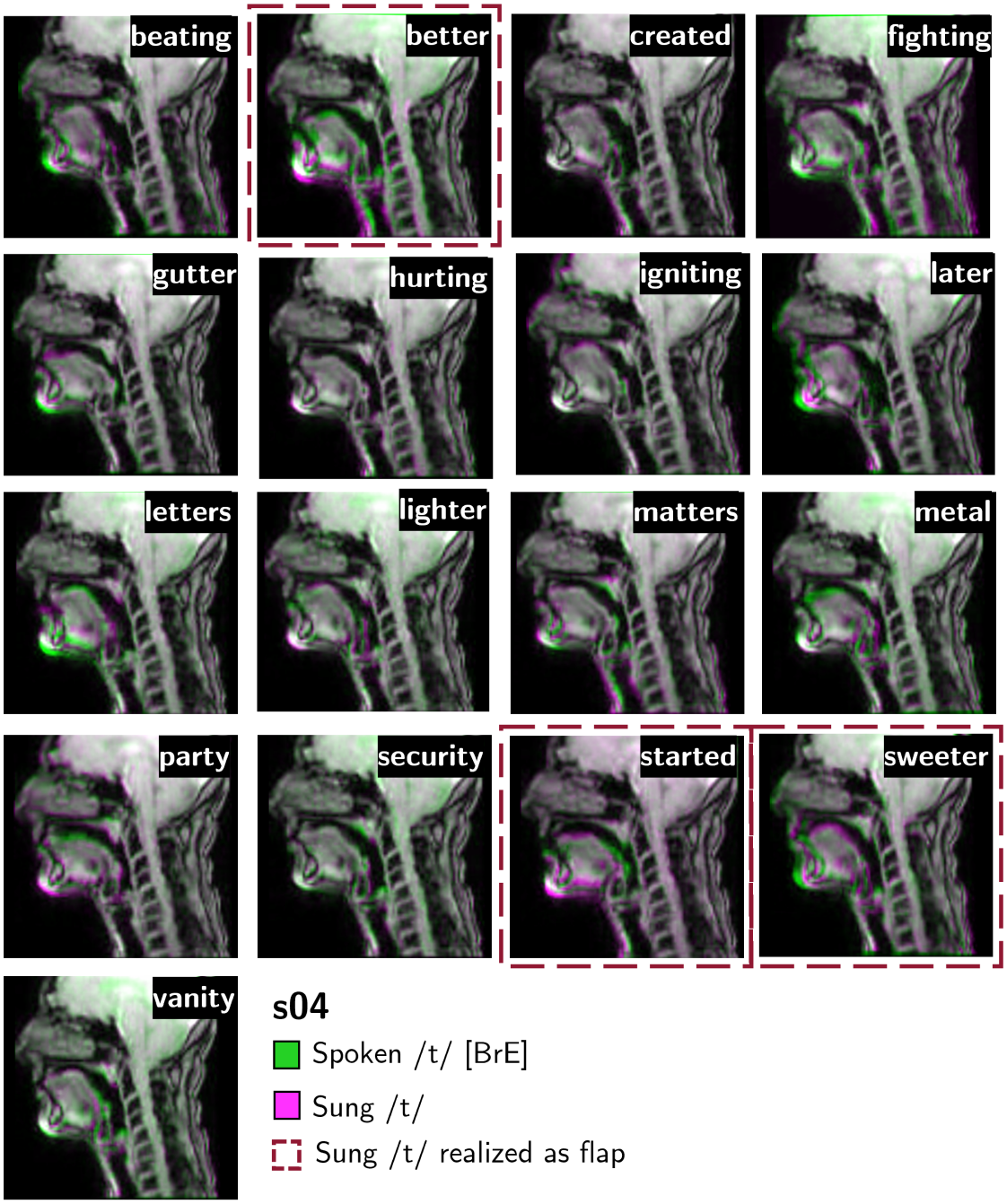


Figure 11.4: *Overlay of MRI frames of spoken /t/ (green) and sung /t/ (magenta) after registration for participant s04.*

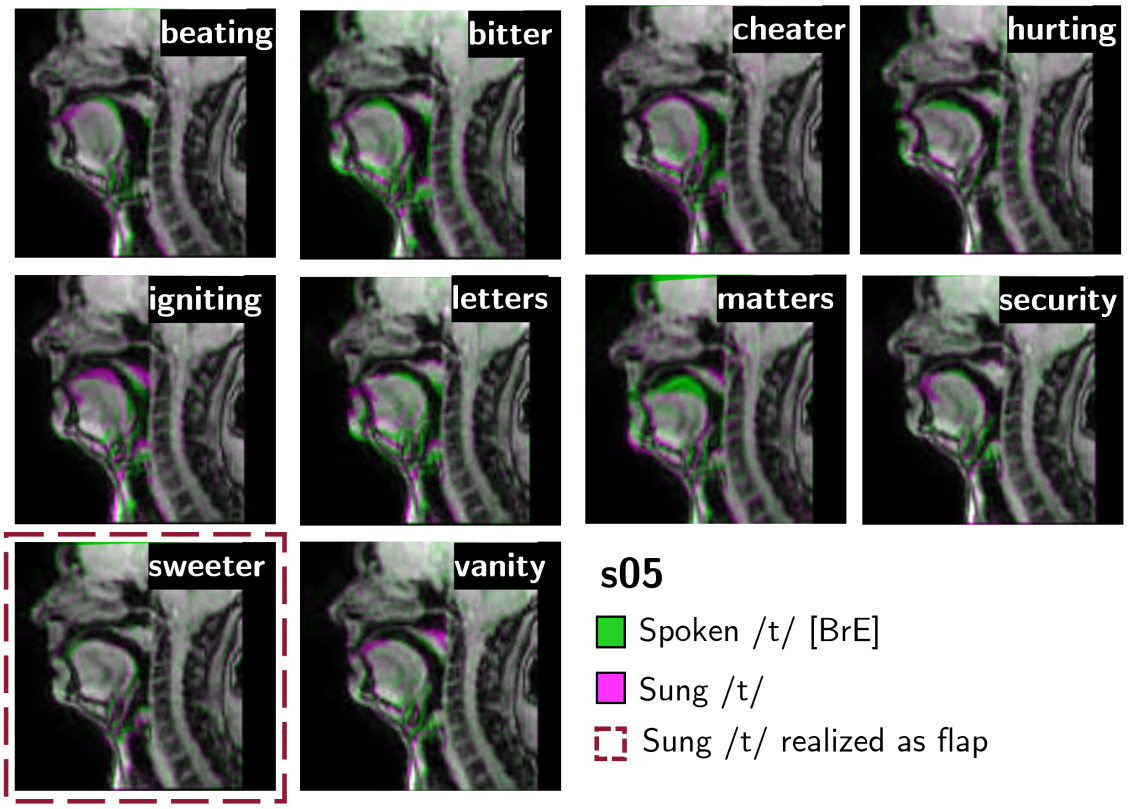


Figure 11.5: *Overlay of MRI frames of spoken /t/ (green) and sung /t/ (magenta) after registration for participant s05.*

Distance estimation

Distance estimation was measured between the tip of the chin and the labial tubercle, resulting in a proxy measurement for jaw opening. These show the same tendency identified in the lip aperture analysis: the distance between the two reference points is greater in the sung condition. In other words, there is greater (inferred) jaw lowering when participants are singing than when they are reading. Table 11.2 displays the average distances in both pixels and normalized per-participant (z-score). A violin

Participant	Spoken (px)	Spoken (norm)	Sung (px)	Sung (norm)
s04	85.3	-0.567	95.4	0.520
s05	75.6	-0.386	78.6	0.225

Table 11.2: *Average distance between labial tubercle and chin tip (in pixels and z-scored) across voice conditions per participant.*

plot, Figure 11.6, was created to visualize the data, where z-score normalized distance distributions in speaking and singing are superimposed and shown by participant.

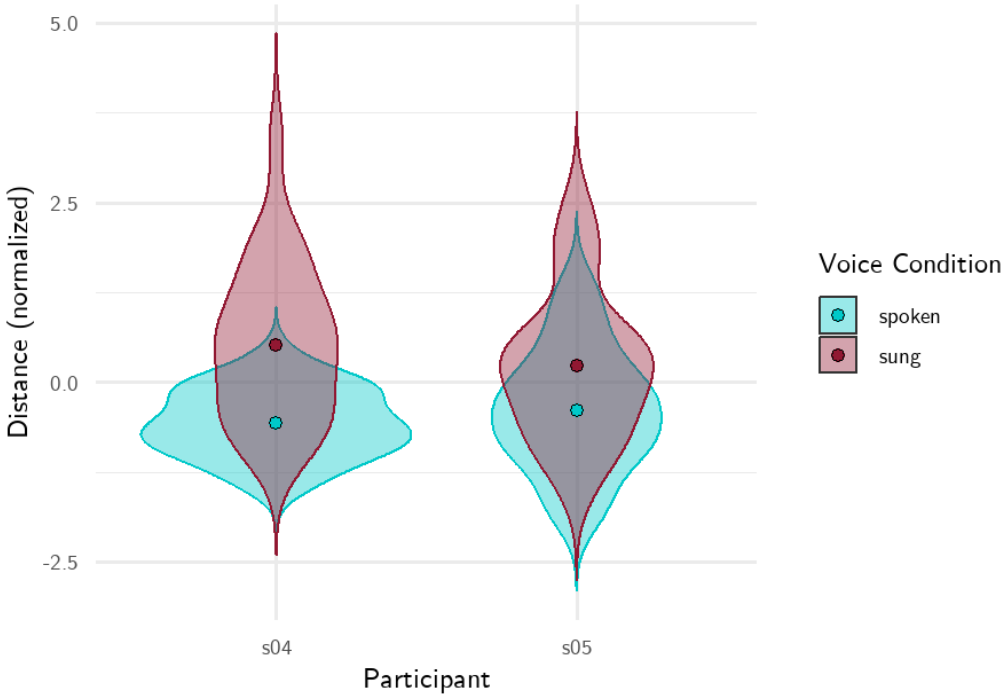


Figure 11.6: *z-score normalized distribution of distances between the chin tip and labial tubercle for participants s04 and s05 as a proxy for jaw opening, across spoken (cyan) and sung (dark red) voice conditions.*

11.4.3 Speaking vs. singing comparison: discussion

The first observation that can be made when looking at Figures 11.4 and 11.5 is that registration was not optimal for some cases. Frames with a slight tint of green or magenta outside of the vocal tract area, notably in the brain or spinal cord (see for example, *fighting* for s04 or *bitter* for s05) indicate position differences in areas that should have been stable landmarks. This may indicate the presence of noise in the target vocal tract section of the overlaid images. Despite clear instructions from both the experimenter and the MRI engineer to keep their heads still, participants found it difficult, especially while reading, and even more so while singing, which naturally encourages tilting the head back and opening the vocal tract significantly.

Nevertheless, some trends can be identified in these images. First, setting aside the tokens realized as flaps and in keeping with results discussed in Production Experiment 2, s04 is overall quite stable and consistent in his sung and spoken /t/ realizations: a majority of frames show very little differences between the two conditions, especially in tongue positioning. For example, frames for *hurting* or *created* in Figure 11.4 are overwhelmingly gray, indicating similar articulatory positioning for the production of intervocalic stop /t/ in both sung and spoken conditions. Participant s05, on the other hand, shows more variation in his spoken and sung /t/ productions, as more color can be seen overall in his images. He also has a more discernible pattern of contrast between spoken and sung tokens, where a majority of spoken tokens (green) are produced with a higher tongue position than sung tokens. The only exception to that is *igniting*, more than likely because the frame did not capture the exact moment of occlusion for the spoken /t/. His pattern corresponds to traditional accounts of singing where a more open vocal tract and therefore lower tongue positioning are expected, as described in Chapter 3.

Participant s05 shows very little contrast in the overlay between spoken and sung /t/ in the flapped token (*sweeter*), in keeping with findings by Fukaya and Byrd (2005) discussed in Section 4.1.2. Flapped tokens for s04 (*better*, *started*, and *sweeter*) seemingly show the biggest differences in articulatory alignment. A closer look at *started* and *sweeter* reveals that perhaps the frames once again did not capture the exact moment of occlusion in both conditions–especially for started where the tongue tip for the spoken /t/ appears nowhere near the alveolar ridge. Although *better* seems to show some misalignment, the contrast in tongue position is interesting, as the spoken stop /t/ seems more laminal than the sung flapped /t/ which looks a lot more apical in its contact with the alveolar ridge.

Regarding distance estimation, results confirm prior research on lip aperture in that singing tends to cause greater jaw lowering on average, as inferred by proxy measures

of the distance between the chin tip and labial tubercle. Figure 11.6 evident variation across the two participants, where s05 exhibits greater overlap between the spoken and sung distributions, suggesting that despite singing leading to greater openings, the difference between the two conditions is less extreme than for s04, who displays greater maximum jaw openings. Looking more closely at each token, Figure 11.7 shows bar plots comparing the sung and spoken distance measurements for each word. This concerned only words that had a sung and spoken condition (hence why s05 has fewer tokens, since some of his spoken data was lost). The raw distance values were kept in Figure 11.7 so the y-axis is not on the exact same scale for s04 than for s05. Across most words, jaw opening was greater in the sung condition compared to the spoken condition, with s04 showing pronounced differences for several words such as *letters*, *vanity*, *shatters*. Words that were flapped by the participants–*better*, *started*, and *sweeter* for s04 and *sweeter* for s05–showed a larger opening in the sung condition than in the spoken condition (for the preceding vowel). However, they do not appear to match a specific trend where a greater opening would more likely lead to a flap realization. Since distances were measured on the frame of maximum opening of the vowel preceding the intervocalic /t/ token, vowel quality likely had an impact on the data, as the jaw has been found to play an active role in influencing vowel height in speech (Mooshammer et al., 2007). Tokens were organized in three categories (open, mid, and close) according to vowel quality, the details of which can be found in Table 11.3. Boxplots showing the distribution of data across vowel quality and participants are displayed in Figure 11.8.

Close	Mid	Open
fighting; igniting; lighter; matters; party; shatters; started	beating; bitter; cheater; created; later; sweeter	better; gutter; hurting; letters; metal; petals; security; vanity; photos

Table 11.3: *Token groups by vowel quality.*

As expected, vowel quality influences jaw opening across both voice conditions to different degrees for each participant. For s04, jaw opening tends to increase with vowel openness in both conditions, with higher overall values for sung tokens. In contrast, s05 shows little difference between voice conditions for mid vowels, but exhibits smaller distances for close vowels and larger distances for open vowels in both conditions, with a noticeable increase in singing. This is in keeping with results from the alignment analysis above. Despite individual variation, these results confirm the hypothesis that singing implies a more open and unobstructed vocal tract, and that part of this result is achieved through jaw lowering, though it did not seem to impact subsequent /t/

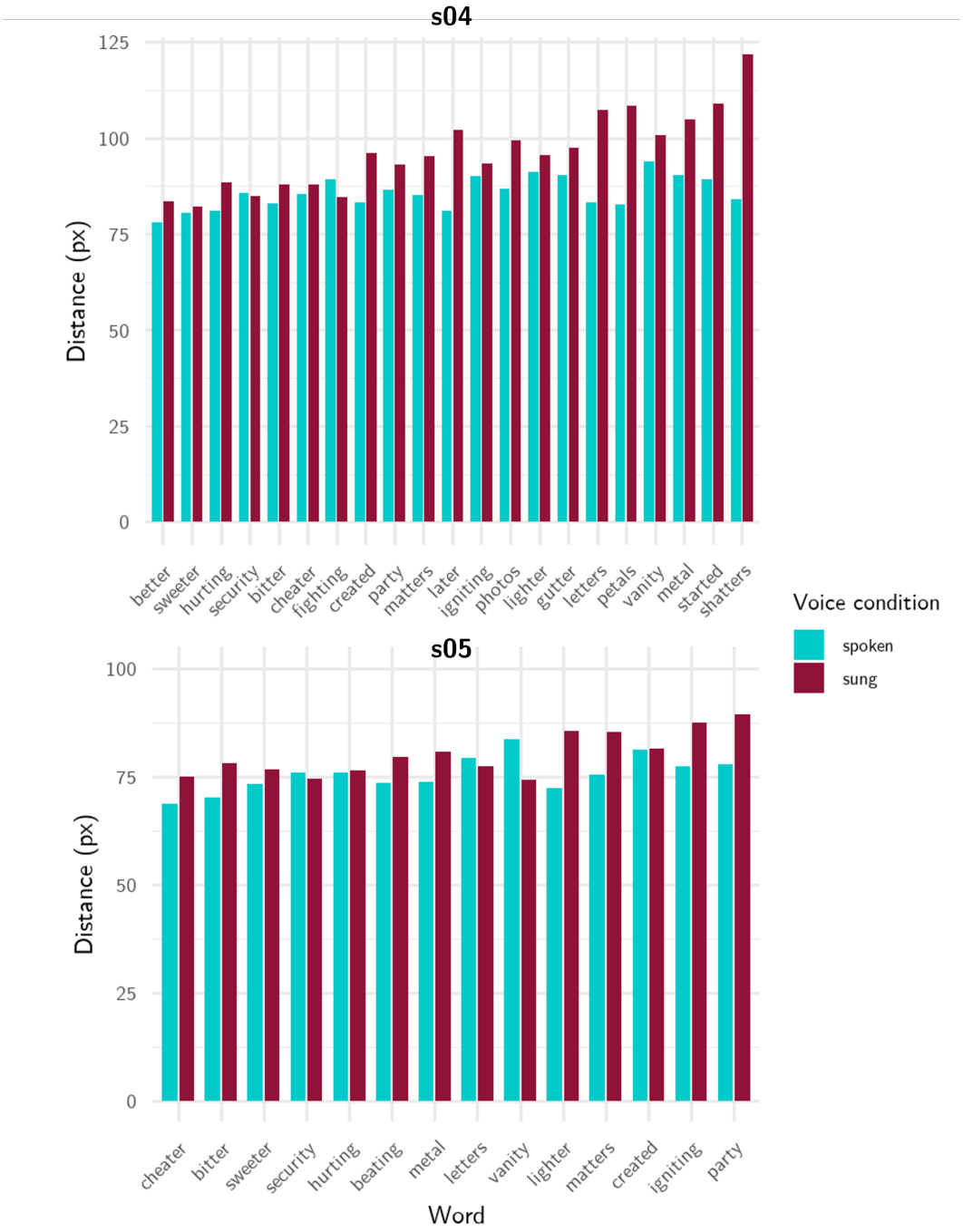


Figure 11.7: *Proxy jaw opening distances (in pixels) for each word, by voice condition, per participant.*

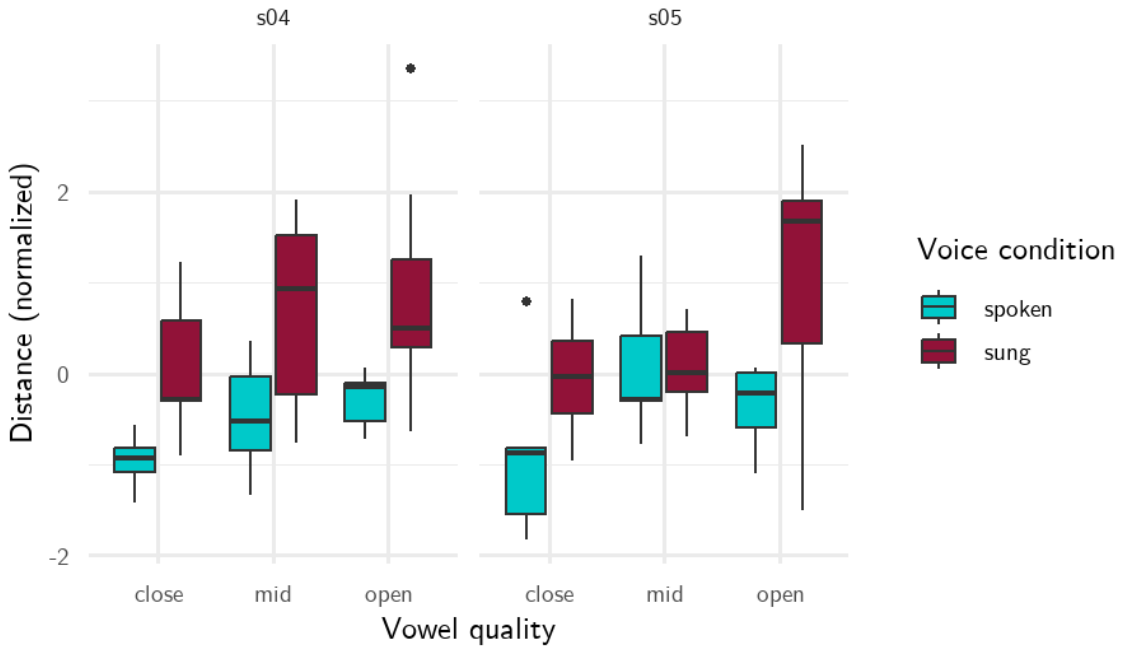


Figure 11.8: *Z-scored proxy jaw opening distance by vowel quality (close, mid, open) and voice condition (spoken vs. sung), for s04 and s05.*

realizations.

11.4.4 A closer look at one occurrence of flapping

Step four in the experiment, READ [AmE], consisted in the participants reading the same sentence they had to sing in what the experimenter said was their "best impression of an American accent." Although obviously read by non-native speakers of this variety of English, and therefore with no guarantee of 'authentic' articulation, this condition was intended to compare potential sung flaps that would be elicited to what the participants assimilated in their mind as a typical American accent. This can be considered in light of Le Page and Tabouret-Keller (1985)'s fourth rider of the Acts of Identity theory, whereby stylistic variation as an act of identity is constrained by the ability of the speaker to actually modify their linguistic behavior. If the participants were able to faithfully imitate an American accent (at least for intervocalic /t/) and flapped when singing, perhaps /t/ is more influenced by style than by articulatory constraints. Conversely, if they had more difficulty with the American accent, yet still flapped when singing, then it could have been argued that articulatory considerations were more influential in this case. However, not enough flap /t/ were elicited when singing to draw any definitive conclusions. Due to the small number of flaps elicited during the experiment, and the fact that the exact moment of constriction was rarely

identifiable on the MRI frames, there is only one token where a sung flap and a read flap can be compared; in the word *sweeter* as sung and read by participant s04. The overlay of the flapped intervocalic /t/ in singing (magenta) and read with an American accent (green) is shown in Figure 11.9. Though the frames are aligned after registration, they

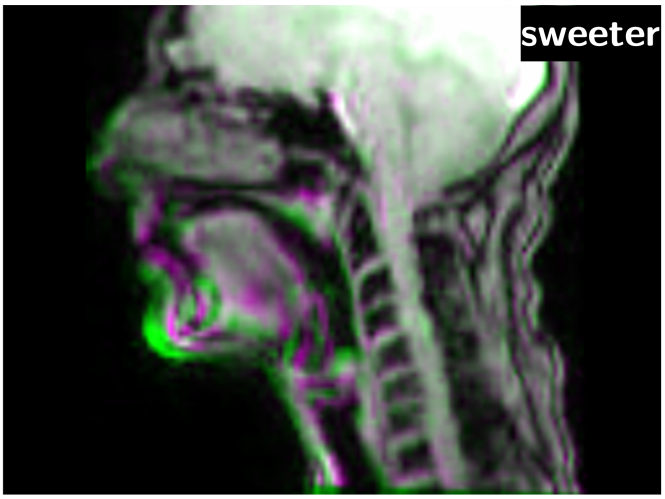


Figure 11.9: *Overlay of MRI frames showing the realization of intervocalic /t/ in* sweeter *as a flap while singing (magenta) and reading with a simulated American accent (green) for participant s04, after registration.*

are not perfectly matched in terms of the exact moment of maximum occlusion. This makes a direct comparison of tongue contact limited, as it appears the tongue is already retracting in the singing flap. Regardless of the tongue tip, the actual positioning of the tongue is very similar in both conditions. The main difference shown in the image is that the head is more tilted back and slightly upwards for singing, consistent with a generally more open vocal tract configuration during singing.

11.5 Chapter summary and conclusion

This chapter provided an exploratory look into the use of MRI for analyzing singing voice pronunciation, with a focus on the production of intervocalic /t/. MRI offers a unique advantage in articulatory research by allowing the visualization of the entire vocal tract. Data was collected for two popular music singers in both singing and reading tasks. Though few flapped tokens were actually elicited, comparing articulatory alignment during the (approximate) occlusion phase of intervocalic stop /t/ did not reveal any extreme of generalizable differences between the two voice conditions. Distance measurements between anatomical landmarks, used as a proxy for jaw opening, indicated greater jaw lowering in the sung condition, broadly confirming results from previous lip tracking analyses, and supporting the first half of Hypothesis 10 whereby

jaw lowering is on average greater in singing than in speech. Nevertheless, an important methodological limitation must be acknowledged in that the 10 Hz temporal resolution (10 images per second) of the MRI scanner was insufficient to reliably capture the exact moment of /t/ occlusion. This resulted in less data than expected, and likely masked fine-grained articulatory differences between voice conditions. Though the current temporal resolution was not ideal for such quick articulations, this chapter still hints at trends that would be interesting to examine in more detail at higher resolutions, and with more participants.

Part III: Summary

Section III consisted in taking a novel articulatory and experimental approach to account for the modification of singing voice pronunciation, with the participation of five British popular music singers. It focused on various modalities shaping the singing voice and tested whether more extreme conditions in terms of buccal aperture, intensity, pitch, and syllable duration had an effect on intervocalic /t/ realization. A concise summary of the main points of each chapter follows:

Chapter 9

This chapter presented the first production experiment of the thesis, seeking to confirm the role of lip aperture in singing as opposed to speech, as well as the influence of intensity and pitch. The results largely confirmed the hypotheses that were tested, showing systematic differences in lip aperture between speech and song and revealing that both pitch and intensity significantly shaped lip aperture patterns, providing a basis for later experiments

Chapter 10

This chapter reported the central experiment of the thesis, which used electropalatography to investigate how articulatory constraints in singing affect the realization of intervocalic /t/. It compared speech and singing, tested effects of syllable duration and pitch, and found that while some trends supported the hypotheses (shorter syllables and higher pitches tended to reduce linguopalatal contact) results were inconsistent and heavily influenced by individual articulatory strategies. Detailed participant analyses revealed substantial inter-speaker variation, with some singers exhibiting flapping and reduced contact, while others maintained stable and prototypical /t/ production.

Chapter 11

This chapter presented an exploratory investigation using MRI to observe vocal tract shaping and jaw dynamics during speech and singing. Data from two singers showed generally greater jaw opening in singing compared to speech, aligning with findings from Chapter 9. While few flapped /t/ tokens were produced and the limited temporal resolution of the scanner restricted fine-grained analysis, some trends were identified that would benefit from further exploration.

General discussion and conclusion

> “If music is a metaphor for identity, ... the self can only be imagined as a particular organization of social, physical, and material forces.”
>
> Frith (2011, p. 109-110)

This thesis has provided a sociocultural and articulatory account of the constraints that govern popular music singing pronunciation. The quotation by sociologist and musicologist Simon Frith touches on the research themes investigated in this work through the study of a single pronunciation feature: a detailed account of the sociophonetics of popular music singing and the role of linguistic choices and articulatory constraints in the construction of a band or singer's identity. These choices are indeed the particular result of the organization of social, physical, and material forces. In other words, together with underlying material considerations, sociocultural and socioindexical forces, as well as physical and articulatory constraints, interact in a dynamic process that ultimately shapes singing voice pronunciation. This general conclusion summarizes the main findings of this thesis by answering the research questions in Section 4.6. It then considers the limitations of the study before discussing future directions this research may take.

11.6 Main findings

Each chapter has addressed its specific hypotheses, and the combined findings from these studies offer potential answers to the overarching research questions that motivate this work:

1. What can be learned from the different genres of the 1980s, as a decade of change in popular music, in terms of the tendency for British singers to style-shift?

The 1980s were a period of transition for the music industry. Pronunciation in popular music evolved from an earlier initiative style-shift toward an American model, to a responsive style-shift, largely unconscious, and responding to a growing standard in popular music to use its own variety of English, known as Pop Song English.

a) Do genres with contrasting sound and image such as traditional heavy metal and pop music both show this tendency? To what extent?

Yes. The distinctively British origins of traditional heavy metal do not seem to supersede the tendency for artists to modify their pronunciation, showing overall very similar flapping tendencies to more mainstream pop music artists. Increased articulatory pressures due to the more extreme singing conditions of this genre may have outweighed its attachment to Britishness. The findings in Chapter 5 and Chapter 6 further support Trudgill (1983, p. 143)'s claim that "there can be no doubt that British singers are modifying their linguistic behaviour for the purposes of singing." That being said, this general trend does not account for the widely variable results when looking at specific artists–the tendency is a highly individual one. This, in part, leads us to the next question:

b) How do respective artists navigate the construction of their artistic identities and genre conventions?

Each artist presents different ways of navigating their identities as well as their attitudes to genre conventions. Whether they are conforming to or rejecting either an overt American model or a more subtle pronunciation standard in popular music, whether accepting the conventions of their subgenres or trying to set themselves apart, British artists consistently had to position themselves linguistically within or against a framework of stylistic norms. This can lead them to perform 'acts of conflicting identity' (Trudgill, 1983).

c) In addition to intervocalic /t/, do other acoustic features also corroborate this tendency?

Yes. In spite of concerns that acoustic analysis would be difficult to achieve with the necessary precision, results have in fact shown that this is possible, and what's more, that the American-inspired pronunciation model of both heavy metal and pop artists of the 1980s not only applies to word-internal intervocalic /t/, but also to vowels such as BATH and LOT.

d) From a broader perspective, is the stop vs. flap allophonic distinction of intervocalic /t/ still as relevant today as it was in the 1980s?

Yes. Despite hinting at a possible change in the coming years, the online identification perception experiment described in Chapter 8, conducted on word-internal intervocalic /t/ in speech, points to the fact that there has not been any relevant evolution in the British population's perception of this feature as American, and that its integration into British English remains limited.

2) Beyond sociolinguistics, and based on the phonetic features of intervocalic /t/ variants, how may articulatory constraints inherent to the singing voice contribute to this modification of pronunciation by British artists? In other words, can a flap /t/ or a flap-like realization occur naturally in singing from the dynamics of the motor system looking to minimize effort, rather than as a distinct indexical goal?

It is difficult to answer this question decisively. While some results provide support in favor of the general hypothesis, that the more extreme or difficult the singing context, the more likely a flap or flap-like realization of word-internal intervocalic /t/ is to occur, these findings are not robust enough to settle the question. Therefore, a more nuanced approach is required, leading to the sub-questions meant to complement the high-level research question:

a) To what extent is mouth opening an important factor in the difference between singing and speaking?

Results from the lip tracking and MRI in Chapter 9 and Chapter 11 studies show that buccal aperture, as studied through proxy measurements of lip aperture and jaw lowering displays significant differences between speech and singing. It was hypothesized that this would therefore have an impact on the realization of a prototypical stop /t/ occlusion, leading to the next question:

b) Do certain musical factors (pitch, syllable duration) dictate which variant of intervocalic /t/ is produced by contemporary popular music British singers?

This does not seem to be the case. Throughout the production experiments, a very small number of auditorily perceptible flaps were elicited. In and of itself, it remains of interest that some did occur in participants who do not exhibit them in speech. Yet, no significant link was found between the manipulated experimental conditions of pitch, intensity, and syllable duration and the realization of intervocalic /t/ as a perceptible flap.

c) How is the articulatory realization of sung intervocalic /t/ impacted in terms of linguopalatal contact and vocal tract configuration?

There are obvious dangers in generalizing from a study of such limited size. However, occasional evidence from the EPG data covered in Chapter 10 suggests that shorter syllable duration and higher pitch may lead to reduced linguopalatal contact and therefore less clearly defined occlusions, despite no perceptible flapping. The MRI study in Chapter 11 found no clear evidence of differences in vocal tract configuration between singing and speaking, though the singing condition showed certain articulatory trends of a more open vocal tract and a lower tongue position. It is possible that experiment issues masked these effects.

3) How do these articulatory and sociocultural constraints work together or against one another to shape what singers ultimately pronounce when singing?

There is no doubt that sociocultural constraints shape, in one way or another, British artists' singing pronunciation. To what extent this resulting American-influenced pronunciation (whether responsive or initiative) is also influenced by the dynamics of the speech/singing motor system remains unclear. Overall, results of the thesis show signs of some articulatory influence, especially in contexts that require more effort, but more data and more studies investigating the question with different experimental designs are needed before this can be claimed robustly.

Whatever the contribution of each constraint on the final pronunciation result, the main constant of this work across both sociocultural and articulatory domains is individual variation. Each distinct artist's pronunciation is shaped by a multitude of factors: from audience influence, identity construction, differentiation and projection, stance-taking in relation to norms and standards, as well as unique articulatory strategies and singing style, down to the influence of a single word. Although a primarily exploratory and descriptive study cannot lead to generalizable claims, it nevertheless is able to highlight how prevalent individual variation is in speech–and in this case, in singing too–at least in genres where, unlike classical singing, no uniform training is provided. Individual characteristics cannot be merely taken as 'noise' to be dismissed from statistical trends and results, instead, accounting for this variation is essential for a holistic understanding of singing voice pronunciation. Working directly with artists provided valuable insight into how these individual differences may manifest themselves, allowing for a richer interpretation overall. The radar chart in Figure 11.10 provides a hypothetical visualization of the relative contribution of articulatory and sociocultural factors for a single artist. It depicts the overall 'shape' that may represent a British popular music singer's unique profile. Factors which have more influence are positioned closer to the outer edge and less prominent ones are closer to the center.

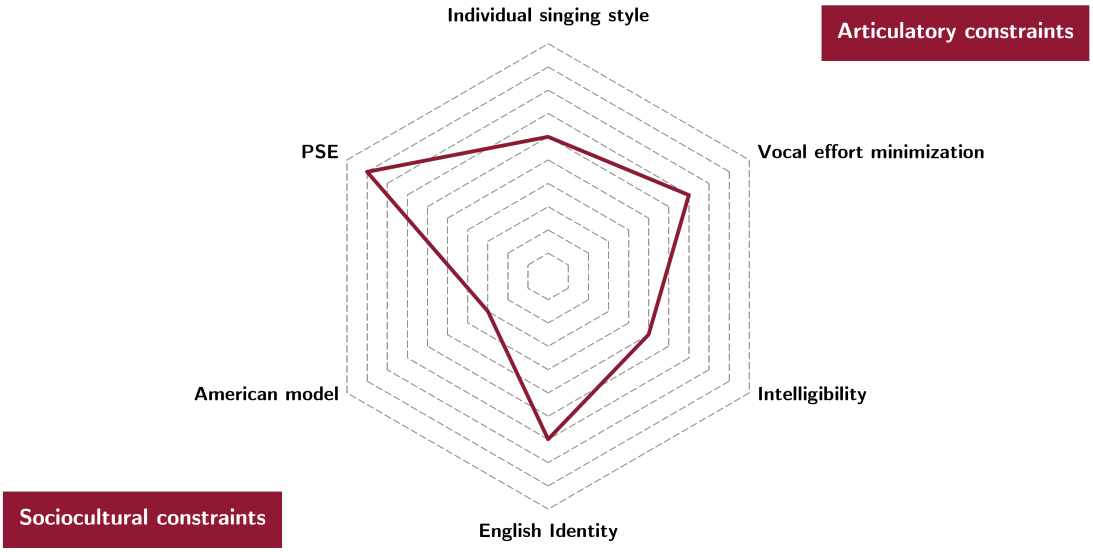


Figure 11.10: *Hypothetical relative contribution of articulatory and sociocultural factors in the sung production of a British popular music artist. Intelligibility was included as a factor for potential further research on the subject, discussed in Section 11.8.*

11.7 Contributions & Implications

This thesis sought to contribute to the body of work on the sociophonetics of the singing voice by examining pronunciation patterns in two genres, heavy metal and pop, positioning them within the specific context of the 1980s, and considering a potential diachronic evolution of the perception of intervocalic /t/ as a Labovian linguistic variable. It aimed to fill an important gap in the field, taking an articulatory phonetics based approach to a pronunciation phenomenon never before studied in depth from this experimental angle. This work supplied additional sociolinguistic, articulatory, perceptual and imaging data related to the flap realization of word-internal intervocalic /t/, which may prove useful in further examining this specific and still elusive allophone. On the methodological level, this work developed a set of novel, albeit perfectible experimental techniques for studying the singing voice, with purpose-designed stimuli, programs, and materials created by the experimenter. The study introduced the possibility for experimenters working with electropalatography to benefit from more accessible opportunities for pseudopalate creation in France. This provides cost-efficient and accessible solutions for future empirical research into the singing voice.

The implications of this thesis are both practical and theoretical. First, the acoustic study in Chapter 7 demonstrated that phonetic signals appear robust enough to be an-

alyzed even in contexts of popular music recordings, where mixed instrumental tracks and added vocal effects could have introduced complexity. This opens up opportunities for more studies in the sociophonetics of popular music singing, not necessarily constrained to consonantal sounds or pre-existing vocals-only tracks. The brevity of the intervocalic flap is one of its main characteristics, and as seen in both MRI and EPG experiments, imaging techniques do not always manage to capture the articulatory gestures of its realization. This was the case for the MRI, limited by the 10 Hz temporal resolution, but the same issue arose with electropalatography as well, where it seemed that even at 200 Hz, capturing a flap proved difficult. Setting aside the technical issues these methods involve, this difficulty to capture the articulatory movement of a flap (across both singing and speaking conditions) might also have wider implications in its categorization. As shown by Warner and Tucker (2011), highly reduced productions of flaps and stops occur very often in American English, in both casual and careful speech–to the point where a majority of them are realized as approximant-like tokens. It is possible therefore that considering a flap as a very rapid stop with a brief closure does not encompass the extent of its articulatory complexity, and that it may be better interpreted as a vowel-like approximant consonant, where the tongue comes very close to the alveolar ridge but does not touch it enough to disrupt the airflow. This in turn could have implications in the sonority perspective put forward by Morrissey (2008), as some approximants closely resemble vowels in their acoustic and articulatory properties. A clearer picture may emerge by reframing our articulatory understanding of flapping and further investigating its acoustic characteristics.

Further theoretical implications of this work involve the perception of flap /t/ from a sound change perspective. Despite some results finding that it may be an increasingly common occurrence among a subset of British youths, mainly for socioindexical reasons (Alderton, 2022), it does not seem to have reached the level of dissemination in other varieties of English as predicted by Wells (1982a). Flapping does not yet seem to be–at least perceptually–a natural or authentic variant of British English. It may therefore be interesting to further investigate this potential phonetic evolution in the upcoming years and give this study value as a record of diachronic change, or its absence, in British English phonetics.

11.8 Limitations and future directions

Limitations pertaining to specific aspects of this thesis have been identified in the relevant sections, but a more general overview is provided here. To start with, the aim of this work was to provide a holistic approach to a multi-faceted phenomenon.

The inevitable limitation of this holistic approach is that while covering a broad range of possible influences, their analysis is not as in-depth or complete as could be found in a work focusing only on one aspect. This may specifically be the case for Part II, where general sociocultural influences and trends for a few artists and bands have been outlined, but providing more detail would have been outside the scope of this thesis. Furthermore, while focusing on a single pronunciation trait permitted this study to work within its necessary constraints, a single trait–no matter how indexical it is–simply cannot provide the full picture of a complex phenomenon such as singing voice pronunciation.

Another limitation of this work is part and parcel of its experimental approach and use of novel techniques: availability and variable quality of the data, technical issues and limitations of the different methods used, and the inevitable pitfalls of the human element in research. These constraints necessitate more data and participants to reach robust conclusions. The combined lip tracking and electropalatography experiment itself may also have been too complex for participants to navigate through its different requirements, and perhaps may have placed excessive demands on non-classically trained singers. Singers themselves had varying ability levels: some were more 'professional' than others, some more talented or at ease in the experimental setting. They were all popular music singers within the broader genre of rock, but belonged to different subgenres and therefore had different approaches to singing, which added a further element of variation.

Lastly, though attempts were made to have as authentic a setting as possible for singers, singing with a pseudopalate in the mouth or inside an MRI scanner does not replicate 'natural' singing conditions. Since no laboratory setup can fully replicate natural singing conditions, future research could explore innovative approaches to find a middle-ground between experimental settings and real-life performance contexts. Data collection could be considered during an artist's studio recording sessions, and adding a comparative element with in-situ live performance recording could prove revealing about direct or indirect audience influence. Another direction could be to use an open upright MRI scanner for data collection, which would considerably increase singers' comfort acquisition and lead to better results. Inevitably, obtaining more data and increasing the number of participants would also help improve the reliability of findings, or conversely, focusing only on participants in targeted case studies could shed light on individual strategies–even more so if these singers are widely recognized to have influenced popular music in one way or another–and provide a starting point for later multi-subject experiments. Future research could extend the sociocultural analysis by including a wider range of artists, genres, and linguistic backgrounds to see if similar

articulatory patterns exist in other contexts, and how these phenomena manifest in diverse musical traditions. Articulatory constraints should apply across languages, so looking at strategies in other languages to reduce vocal effort could offer valuable comparison.

A main aspect that also influences pronunciation in singing is intelligibility. Some discussion of intelligibility took place in the theoretical background as well as when analyzing specific tokens in Section 10.4.2 and Section 11.4.4. However, this facet simply could not be discussed fully within the constraints of this thesis. Sung material is known to involve a considerable loss of intelligibility (Collister & Huron, 2008). Lindblom (1990)'s 'Hyper-' and 'Hypo-' articulation Theory as well as the Smooth Signal Redundancy Hypothesis by Aylett and Turk (2004) briefly described in Chapter 3, highlights the role of intelligibility constraints and offer a strong theoretical foundation for future research from this perspective. Considering a flap as a lenited allophone of /t/, it would be of relevance to study flapping behaviors in singing according to the lyrics themselves, the genre-dependent importance of the text, the lexical frequency of words containing tokens realized as flaps and stops, or where flaps occur within the song (repeated multiple times in the chorus or once in a verse).

Additionally, based on the pilot perception study described in Chapter 8, my next research efforts will focus on improving and replicating this experiment. This will involve correcting the presentation of the stimuli, and incorporating musical material in the vein of Gibson and Hay (2025) to include a singing pronunciation perspective to the experiment. Developing the sociolinguistic questionnaire part of the study will allow for a more comprehensive and clearer assessment of the perception of intervocalic /t/ realizations, as well as a potential sound change evaluation. The singing component will furthermore explore how differently British speakers may perceive flapping in spoken and sung contexts, potentially reflecting the same tension between flap /t/ as a typically American pronunciation trait in speech and a supra-local norm in singing. Furthermore, the MRI data collected in this project could benefit from further analysis. Dynamic methods of image registration may improve the quality of the articulatory alignment analysis: while the rigid registration technique used aligns images using only rotation and translation, dynamic registration methods take into account the frame sequences of an MRI video and therefore improve the alignment of anatomical landmarks. A closer look at automated methods such as the open-source vocal tract shape measurement toolbox proposed by Belyk et al. (2024) could improve the handmade distance estimation analysis. All of these refinements may allow for more reliable data collection.

11.9 Conclusion

To return to the line from "Hey Jude" by The Beatles, cited at the very beginning of this work, Paul McCartney's realization of *better* as [beɾə] aligned with the performed pronunciation of the singers who contributed to the experimental aspect of this work: it was, incidentally, the only word that was realized as a flap by a majority of them. The exact reason for the use of this variant remains unclear, and it is likely competing influences contribute to /t/ flapping. The singing voice is shaped by a complex interplay of linguistic, sociocultural, physiological, and performance-related factors that resist any single, definitive explanation. Only by considering these elements individually, and through their interdependence, much like I show in Figure 11.10 can we move closer to a full description of flapping in British pop music, and what this single phonetic feature reveals about the interrelationships between language, music, and society.

Appendices

A Part II Appendix

A.1 Chapter 5: Heavy metal pronunciation

A.1.1 Albums included in the corpus

Black Sabbath

- *Black Sabbath* (1970)
- *Paranoid* (1970)
- *Master of Reality* (1971)
- *Vol. 4* (1972)
- *Sabbath Bloody Sabbath* (1973)
- *Sabotage* (1975)
- *Technical Ecstasy* (1976)
- *Never Say Die!* (1978)
- *Heaven and Hell* (1980)
- *Mob Rules* (1981)
- *Born Again* (1983)
- *Seventh Star* (1986)
- *The Eternal Idol* (1987)
- *Headless Cross* (1989)
- *Tyr* (1990)
- *Dehumanizer* (1992)
- *Cross Purposes* (1994)
- *Forbidden* (1996)
- *13* (2013)

Judas Priest

- *Rocka Rolla* (1974)
- *Sad Wings of Destiny* (1976)
- *Sin After Sin* (1977)
- *Stained Class* (1978)
- *Killing Machine* (released in the US as *Hell Bent for Leather*) (1978)
- *British Steel* (1980)
- *Point of Entry* (1981)
- *Screaming for Vengeance* (1982)
- *Defenders of the Faith* (1984)

- *Turbo* (1986)
- *Ram it Down* (1988)
- *Painkiller* (1990)
- *Jugulator* (1997)
- *Demolition* (2001)
- *Angel of Retribution* (2005)
- *Nostradamus* (2008)
- *Redeemer of Souls* (2014)
- *Firepower* (2018)

Iron Maiden

- *Iron Maiden* (1980)
- *Killers* (1981)
- *The Number of the Beast* (1982)
- *Piece of Mind* (1983)
- *Powerslave* (1984)
- *Somewhere in Time* (1986)
- *Seventh Son of a Seventh Son* (1988)
- *No Prayer for the Dying* (1990)
- *Fear of the Dark* (1992)
- *The X Factor* (1995)
- *Virtual XI* (1998)
- *Brave New World* (2000)
- *Dance of Death* (2003)
- *A Matter of Life and Death* (2006)
- *The Final Frontier* (2010)
- *The Book of Souls* (2015)

Def Leppard

- *On Through the Night* (1980)
- *High 'n' Dry* (1981)
- *Pyromania* (1983)
- *Hysteria* (1987)
- *Adrenalize* (1992)
- *Slang* (1996)
- *Euphoria* (1999)
- *X* (2002)
- *Yeah!* (2006)
- *Songs from the Sparkle Lounge* (2008)
- *Def Leppard* (2015)

A.1.2 Albums by American bands included in the corpus

Mötley Crüe

- *Too Fast for Love* (1981)
- *Shout at the Devil* (1983)
- *Theatre of Pain* (1985)
- *Girls, Girls, Girls* (1987)
- *Dr. Feelgood* (1989)

Ratt

- *Out of the Cellar* (1984)
- *Invasion of your Privacy* (1985)
- *Dancing Undercover* (1986)
- *Reach for the Sky* (1988)

A.1.3 Interviews

Black Sabbath

Interview 1 – 1970: The description at the beginning of the video states that the band was supposed to take part in an anti-American demonstration in Germany, but withdrew after seeing the flags of Vietnam, Cambodia and Greece.
Last Accessed 04/09/2025 - https://www.youtube.com/watch?v=Vzl_e9uLNHA

Interview 2 – 1976: Radio program Innerview on the L.A. station KMET. Interview conducted by Jim Ladd.
Last Accessed 13/05/2021 - No longer available on YouTube.

Interview 3 – 1982: Ozzy Osbourne on the American TV program Late Night with David Letterman.
Last Accessed 04/09/2025 - https://www.youtube.com/watch?v=pfTPr0eG8Og

Interview 4 – 1987: Tony Martin and Tony Iommi on British TV program Monsters of Rock, hosted by Mick Wall on Sky TV.
Last Accessed 04/09/2025 – https://www.youtube.com/watch?v=adRPYxVBw8I

Interview 5 – 1991: Tonny Iommi on the band's documentary The Black Sabbath Story Vol. 1 - 1970-1978. Interviewer unknown.
Last Accessed 13/05/2021 – No longer available on YouTube.

Interview 6 – 2017: BBC News Interview of Ozzy Osbourne. Interviewer unknown.
Last Accessed 04/09/2025 – https://www.youtube.com/watch?v=0IM5b6rvfpU&t=106s

Judas Priest

Interview 1 – 1977: Ian "Molly" Meldrum interview of Rob Halford and Glenn Tipton about their new album *Sin After Sin* on Australian TV program *Countdown*.
Last Accessed 04/09/2025 – https://www.youtube.com/watch?v=JBhuKOuIUCo

Interview 2 – 1984: MTV interview of Rob Halford for the release of *Defenders of the Faith* on *The Week in Rock* program by Mark Goodman.
Last Accessed 04/09/2025 – https://www.youtube.com/watch?v=PzvZIYQU3Os

Interview 3 – 1986: Rob Halford interview with Fred Graham on CBS News Nightwatch.
Last Accessed 04/09/2025 – https://www.youtube.com/watch?v=YDlGFvhLI-0

Interview 4 – 1988: Rob Halford and K. K. Downing on British TV program *Monsters of Rock*, hosted by Mick Wall on Sky TV.
Last Accessed 04/09/2025 – https://www.youtube.com/watch?v=dqnsHbon7ec

Interview 5 – 1990: Rob Halford and K. K. Downing on American TV (Channel and interviewer unknown).
Last Accessed 04/09/2025 – https://www.youtube.com/watch?v=_UARHsYZcko

Interview 6 – 2003: *Rockthology Vol. 3: Legends of Rock* interview of Rob Halford in his home in Phoenix, Arizona.
Last accessed 04/09/2025 – https://www.youtube.com/watch?v=XFdsLkrSiT8

Interview 7 – 2004: Interview of the whole band for the *Judas Priest Reunited Tour* Documentary.
Last Accessed: 04/09/2025 – https://www.youtube.com/watch?v=JR7EfBba2l4

Interview 8 – 2018: Rob Halford and Richie Faulkner interview on New York radio station Q1043 conducted by Jonathan Clarke.
Last Accessed: 04/09/2025 – https://www.youtube.com/watch?v=jPRBG_AQC8E

Iron Maiden

Interview 1 – 1982: Steve Harris, Clive Burr and Bruce Dickinson on British kid's TV program *Tiswas*. Interviewer unknown.
Last Accessed 13/05/2021 – No longer available on YouTube.

Interview 2 – 1985: Bruce Dickison on British TV program *Monsters of Rock*, special New Year's Eve episode hosted by Mick Wall and unknown co-host on Sky TV.
Last Accessed: 04/09/2025 – https://www.youtube.com/watch?v=VzEOhivDfRQ

Interview 3 – 1988: Mick Wall's *Monsters of Rock interview* of the band in Germany after having finished recording *Seventh Son of a Seventh Son.*
Last Accessed 13/05/2021 – No longer available on YouTube.

Interview 4 – 1988: Bruce Dickinson and Dave Murray interviewed by Erika Ehm on Canadian TV program *MuchMusic*.
Last Accessed: 04/09/2025 – https://www.youtube.com/watch?v=AYrzR_4QBrs

Interview 5 – 1992: Bruce Dickinson interview by John Norris on MTV.
Last Accessed 04/09/2025 – https://www.youtube.com/watch?v=8hys_zhfvZU

Interview 6 – 1993: MTV's *Headbanger's Ball* Iron Maiden special episode, hosted by Vanessa Warwick.
Last Accessed 04/09/2025 – https://www.youtube.com/watch?v=PFwH-wDC-QU

Interview 7 – 1999: Rock Radio Network on VH-1 interview of Iron Maiden by Tommy Vance.
Last Accessed: 04/09/2025 – https://www.youtube.com/watch?v=ymDAjOYCHtE

Interview 8 – 2004: Bruce Dickinson interviewed by Sam Dunn on the Hammersmith stage for the documentary *Metal: A Headbanger's Journey*.
Last Accessed 04/09/2025 – https://www.youtube.com/watch?v=SR9n7xAh2v0

Interview 9 – 2010: Iron Maiden interviewed by Sam Dunn for *The Final Frontier* short documentary.
Last Accessed 13/05/2021 – No longer available on YouTube.

Def Leppard

Interview 1 – 1981: Steve Clark and Joe Elliott interview on the American radio program *Rockline* hosted by Bob Coburn.
Last Accessed: 10/09/2025 – https://www.youtube.com/watch?v=rawONf-LOdY

Interview 2 – 1983: Joe Elliott and Rick Savage American TV interview. Interviewer unknown.
Last Accessed 10/09/2025 – https://www.youtube.com/watch?v=66PYErWUQ1g

Interview 3 – 1987: Joe Elliott interviewed by Mick Wall for the *Monsters of Rock* TV program on Sky TV.
Last Accessed on 10/09/2025 – https://www.youtube.com/watch?v=k-13HMCKumo

Interview 4 – 1988: Def Leppard interviewed by kids on Norwegian TV program *Blikkbåx*,

airing on NRK.
Last Accessed 10/09/2025 – https://www.youtube.com/watch?v=KNT9vllK-Qc

Interview 5 – 1992: Rick Savage interview at the Silver Clef Awards, organized by the Nordoff-Robbins Center, a charity that uses music to help children who have mental or physical disabilities to communicate with those around them.
Last Accessed on 10/09/2025 – https://www.youtube.com/watch?v=EvYlhIxgLO4

Interview 6 – 1996: Rick Savage and Joe Elliott interview in Paris studio, aired on French TV channel MCM.
Last Accessed 10/09/2025 – https://www.youtube.com/watch?v=-zG_hIlozo0

Interview 7 – 2000: California State Fair Expo interview of Joe Elliott by "Metal" Matt from C.C. Rock.
Last Accessed 10/09/2025 – https://www.youtube.com/watch?v=iw5-8TV4Dl8

Interview 8 – 2015: Joe Elliott and Rick Savage interview by Mark Douglas for YoutubePresents.
Last Accessed 10/09/2025 – https://www.youtube.com/watch?v=WnNURFIL9Vo

A.2 Chapter 6: Pop music pronunciation

A.2.1 Albums, selected songs, and interviews included in the corpus

ABC

- *The Look of Love* (1982): All of my Heart; Poison Arrow; The Look of Love

Interview 1 (1980-1983); Interview 2 (1984-1986); Interview 3 (1987-1989); Contemporary Interview 1; Contemporary Interview 2; Contemporary Interview 3

Adam and the Ants

- *Kings of the Wild Frontier* (1980): Don't Be Square (Be There); Feed Me to the Lions; Jolly Roger; Kings of the Wild Frontier; Los Rancheros
- *Prince Charming* (1981): Stand and Deliver; That Voodoo!

Interview 1 (1980-1983); Interview 2 (1984-1986); Interview 3 (1987-1989); Contemporary Interview 1; Contemporary Interview 2; Contemporary Interview 3

Brian Ferry

- *Boys and Girls* (1986): Don't Stop the Dance; Sensation

Interview 1 (1980-1983); Interview 2 (1984-1986); Interview 3 (1987-1989);
Contemporary Interview 1; Contemporary Interview 2; Contemporary Interview 3

Bros

- *Push* (1988): Drop the Boy; Liar; Love to Hate You; Shocked; Ten out of Ten; When Will I Be Famous?

NA; Interview 2 (1984-1986); Interview 3 (1987-1989);
Contemporary Interview 1; Contemporary Interview 2; Contemporary Interview 3

Cliff Richard

- *Wired for Sound* (1985): Better Than I Know Myself; Broken Doll; Oh No, Don't Let Go; Summer Rain; Wired for Sound
- *Always Guaranteed* (1987): Once Upon a Time; One Night; Two Hearts
- *Stronger* (1989): Clear Blue Skies; I Just Don't Have the Heart; Joanna; The Best of Me

Interview 1 (1980-1983); Interview 2 (1984-1986); Interview 3 (1987-1989);
Contemporary Interview 1; Contemporary Interview 2; Contemporary Interview 3

Culture Club

- *Colour by Numbers* (1983): Changing Every Day; Church of the Poison Mind; It's a Miracle; Miss Me Blind; Mister Man

Interview 1 (1980-1983); Interview 2 (1984-1986); Interview 3 (1987-1989);
Contemporary Interview 1; Contemporary Interview 2; Contemporary Interview 3

Curiosity Killed the Cat

- *Keep your Distance* (1987): Down to Earth; Free; Ordinary Day

Interview 1 (1980-1983); Interview 2 (1984-1986); Interview 3 (1987-1989);
Contemporary Interview 1; NA; NA

David Bowie

- *Scary Monsters (and Super Creeps)* (1980): Ashes To Ashes; Scary Monsters (and Super Creeps); Scream Like a Baby; Teenage Wildlife; Up The Hill Backwards
- *Let's Dance (1983)*: China Girl; Criminal World; Let's Dance; Shake It

Interview 1 (1980-1983); Interview 2 (1984-1986); Interview 3 (1987-1989);
Contemporary Interview 1; Contemporary Interview 2; Contemporary Interview 3

Depeche Mode

- *Violator* (1990): Blue Dress; Clean; Enjoy the Silence; Sweetest Perfection; World in My Eyes

Interview 1 (1980-1983); Interview 2 (1984-1986); Interview 3 (1987-1989);
Contemporary Interview 1; Contemporary Interview 2; Contemporary Interview 3

Dexys Midnight Runners

- *Too Rye Ay* (1982): I'll Show You; Jackie Wilson Said (I'm In Heaven When You Smile)

Interview 1 (1980-1983); Interview 2 (1984-1986); Interview 3 (1987-1989);
Contemporary Interview 1; Contemporary Interview 2; NA

Duran Duran

- *Duran Duran* (1981): Anyone Out There; Careless Memories; Friends of Mine; Sound of Thunder; To the Shore
- *Rio* (1982): Hungry Like the Wolf; My Own Way; The Chauffeur
- *Seven and the Ragged Tiger* (1983): New Moon on Monday; Of Crime and Passion; The Seventh Stranger; Union of the Snake

Interview 1 (1980-1983); Interview 2 (1984-1986); Interview 3 (1987-1989);
Contemporary Interview 1; Contemporary Interview 2; Contemporary Interview 3

Electric Light Orchestra

- *Time* (1981): From the End of the World; Here is the News; Hold on Tight; The Lights Go Down; Ticket To The Moon; Twilight

Interview 1 (1980-1983); NA; Interview 3 (1987-1989);
Contemporary Interview 1; Contemporary Interview 2; Contemporary Interview 3

Elton John

- *Too Low for Zero* (1983): Cold As Christmas (In the Middle of the Year); I'm Still Standing; One More Arrow; Religion; Saint; Whipping Boy
- *Breaking Hearts* (1984): Breaking Hearts Ain't What It Used To Be; Burning Buildings; Did He Shoot Her?; Lil' 'Frigerator; Restless; Sad Songs (Say So Much)
- *Ice on Fire* (1985): Candy By the Pound; Cry to Heaven; Nikita; Tell Me What the Papers Say; This Town; Wrap Her Up
- *Sleeping with the Past* (1989): Amazes Me; Blue Avenue; Club At the End of the Street; Durban Deep; Healing Hands; Sleeping With the Past; Stone's Throw from Hurtin'

Interview 1 (1980-1983); Interview 2 (1984-1986); Interview 3 (1987-1989);
Contemporary Interview 1; Contemporary Interview 2; Contemporary Interview 3

Erasure

- *The Circus* (1987): Don't Dance; It Doesn't Have To Be
- *The Innocents* (1988): A Little Respect; Hallowed Ground; Imagination; Ship of Fools; Yahoo!
- *Wild* (1989): Blue Savannah; Brother and Sister; Drama!; How Many Times?; Piano Song; Star

Interview 1 (1980-1983); Interview 2 (1984-1986); Interview 3 (1987-1989);
Contemporary Interview 1; Contemporary Interview 2; Contemporary Interview 3

Eric Clapton

- *Journeyman* (1989): Hard Times; Lead Me On; Old Love; Running On Faith

Interview 1 (1980-1983); Interview 2 (1984-1986); Interview 3 (1987-1989);
Contemporary Interview 1; Contemporary Interview 2; Contemporary Interview 3

Frankie Goes to Hollywood

- *Welcome to the Pleasuredome* (1984): Born To Run; Snatch of Fury (Stay); Relax; San Jose (The Way); The Power of Love; War

NA; Interview 2 (1984-1986); Interview 3 (1987-1989);
Contemporary Interview 1; Contemporary Interview 2; Contemporary Interview 3

Genesis

- *Duke* (1980): Alone Tonight; Behind The Lines; Cul de Sac; Duke's Travel; Man Of Our Times; Misunderstanding; Open Door; Turn It On Again
- *Abacab* (1981): Another Record; Keep It Dark; Like It or Not; Me and Sarah Jane
- *Genesis* (1983): Just a Job to Do
- *Invisible Touch* (1986): Anything She Does; In Too Deep; Invisible Touch

Interview 1 (1980-1983); Interview 2 (1984-1986); Interview 3 (1987-1989);
Contemporary Interview 1; Contemporary Interview 2; Contemporary Interview 3

George Michael

- *Faith* (1987): Father Figure; Monkey
- *Listen without Prejudice* (1990): Cowboys and Angels; Freedom! '90; Mothers Pride; Praying Time; They Won't Go When I Go; Waiting for That Day

Interview 1 (1980-1983); Interview 2 (1984-1986); Interview 3 (1987-1989);
Contemporary Interview 1; Contemporary Interview 2; Contemporary Interview 3

Go West

- *Go West* (1985): Call Me; Eye to Eye; Goodbye Girl; Haunted; Missing Persons

Interview 1 (1980-1983); NA; Interview 3 (1987-1989);
Contemporary Interview 1; Contemporary Interview 2; Contemporary Interview 3

Haircut One Hundred

- *Pelican West* (1985): Calling Captain Autumn; Fantastic Day; Kingsize (You're My Little Steam Whistle); Love Plus One; Marine Boy; Milk Film

Interview 1 (1980-1983); Interview 2 (1984-1986); Interview 3 (1987-1989);
Contemporary Interview 1; Contemporary Interview 2; NA

Heaven 17

- *The Luxury Gap* (1983): Crushed By the Wheels of Industry; The Best Kept Secret; Who'll Stop the Rain

Interview 1 (1980-1983); NA; Interview 3 (1987-1989);

Contemporary Interview 1; Contemporary Interview 2; Contemporary Interview 3

Johnny Hates Jazz

- *Turn Back the Clock* (1988): What Other Reason

NA; NA; NA;

Contemporary Interview 1; Contemporary Interview 2; Contemporary Interview 3

Level 42

- *World Machine* (1985): Coup D'Etat; Dream Crazy; It's Not the Same For Us
- *Running in the Family* (1987): Children Say; It's Over; The Sleepwalkers; Two Solitudes (Everyone's Love In The Air)

Interview 1 (1980-1983); Interview 2 (1984-1986); Interview 3 (1987-1989);

Contemporary Interview 1; Contemporary Interview 2; Contemporary Interview 3

Nik Kershaw

- *Human Racing* (1984): Bogart
- *The Riddle* (1984): City of Angels; Wild Horses

Interview 1 (1980-1983); Interview 2 (1984-1986); Interview 3 (1987-1989);

Contemporary Interview 1; Contemporary Interview 2; Contemporary Interview 3

Orchestral Manoeuvres in the Dark

- *Architecture and Morality* (1981): She's Leaving; Souvenir; The New Stone Age

Interview 1 (1980-1983); Interview 2 (1984-1986); Interview 3 (1987-1989);

Contemporary Interview 1; Contemporary Interview 2; Contemporary Interview 3

Paul McCartney

- *Tug of War* (1982): Ballroom Dancing; Dress Me Up as a Robber; Ebony and Ivory; Here Today; Somebody Who Cares; Wanderlust

Interview 1 (1980-1983); Interview 2 (1984-1986); Interview 3 (1987-1989);

Contemporary Interview 1; Contemporary Interview 2; Contemporary Interview 3

Paul Young

- *No Parlez* (1983): Ku Ku Kurama
- *The Secret of Association* (1985): I'm Gonna Tear Your Playhouse Down; Man in the Iron Mask; Soldier's Things

Interview 1 (1980-1983); Interview 2 (1984-1986); Interview 3 (1987-1989);

Contemporary Interview 1; Contemporary Interview 2; Contemporary Interview 3

Pet Shop Boys

- *Please* (1986): I want a Lover; Tonight is Forever; Two Divided by Zero; Violence; West End Girls
- *Actually* (1987): Hit Music; One More Chance; Shopping
- *Introspective* (1988): Always On My Mind / In My House; Domino Dancing; I Want a Dog

Interview 1 (1980-1983); Interview 2 (1984-1986); Interview 3 (1987-1989);
Contemporary Interview 1; Contemporary Interview 2; Contemporary Interview 3

Peter Gabriel

- *So* (1986): Mercy Street; That Voice Again

Interview 1 (1980-1983); Interview 2 (1984-1986); Interview 3 (1987-1989);
Contemporary Interview 1; Contemporary Interview 2; Contemporary Interview 3

Phil Collins

- *Face Value* (1981): I Missed Again; If Leaving Me Is Easy; I'm Not Moving; In the Air Tonight; This Must be Love; Tomorrow Never Knows
- *No Jacket Required* (1983): Don't Lose My Number; Inside Out; One More Night; Who Said I Would
- *...But Seriously* (1989): All of my Life; Another Day in Paradise; Colours; Do You Remember?; Father to Son; Hang in Long Enough; I Wish it Would Rain Down; That's Just the Way It Is

Interview 1 (1980-1983); Interview 2 (1984-1986); Interview 3 (1987-1989);
Contemporary Interview 1; Contemporary Interview 2; Contemporary Interview 3

Pink Floyd

- *The Wall* (1979): Another Brick in The Wall; Comfortably Numb; One of My Turns; The Thin Ice; Waiting For The Worms; Young Lust

Interview 1 (1980-1983); Interview 2 (1984-1986); Interview 3 (1987-1989);
Contemporary Interview 1; Contemporary Interview 2; Contemporary Interview 3

Roxy Music

- *Flesh and Blood* (1980): Oh Yeah!; Over You; Running Wild: Same Old Scene
- *Avalon* (1982): Avalon; More Than This; Take A Chance With Me; To Turn You On

Interview 1 (1980-1983); Interview 2 (1984-1986); Interview 3 (1987-1989);
Contemporary Interview 1; Contemporary Interview 2; Contemporary Interview 3

Simply Red

- *Picture Book* (1985): Jericho; Money's Too Tight (to Mention); No Direction; Open Up the Red Box; Picture Book
- *Men and Women* (1987): Ev'ry Time We Say Goodbye; I Won't Feel Bad; Let Me Have It All; Love Fire; Move On Out; Shine; Suffer; The Right Thing
- *A New Flame* (1989): More; She'll Have to Go

Interview 1 (1980-1983); NA; Interview 3 (1987-1989);
Contemporary Interview 1; Contemporary Interview 2; Contemporary Interview 3

Soft Cell

- *Non Stop Erotic Cabaret* (1982): Bedsitter; Chips On My Shoulder; Entertain

Me; Frustration

Interview 1 (1980-1983); Interview 2 (1984-1986); Interview 3 (1987-1989);
Contemporary Interview 1; Contemporary Interview 2; Contemporary Interview 3

Spandau Ballet

- *True* (1983): Foundation; Heaven Is a Secret; True
- *Parade* (1984): Highly Strung; I'll Fly For You; Only When You Leave

Interview 1 (1980-1983); Interview 2 (1984-1986); Interview 3 (1987-1989);
Contemporary Interview 1; Contemporary Interview 2; Contemporary Interview 3

Sting

- *The Dream of the Blue Turtles* (1985): Consider Me Gone; Love is the Seventh Wave; Russians

Interview 1 (1980-1983); Interview 2 (1984-1986); Interview 3 (1987-1989);
Contemporary Interview 1; Contemporary Interview 2; Contemporary Interview 3

Tears for Fears

- *The Hurting* (1983): Change; Mad World; Start of the Breakdown; The Prisoner; Watch Me Bleed
- *Songs From the Big Chair* (1985): Listen
- *The Seeds of Love* (1989): Swords and Knives

Interview 1 (1980-1983); Interview 2 (1984-1986); Interview 3 (1987-1989);
Contemporary Interview 1; Contemporary Interview 2; Contemporary Interview 3

The Beautiful South

- *Welcome to the Beautiful South* (1989): I'll Sail this Ship Alone; Woman in the Wall
- *Choke* (1990): I've Come For my Award; Let Love Speak Up Itself; My Book

Interview 1 (1980-1983); NA; NA;
Contemporary Interview 1; Contemporary Interview 2; Contemporary Interview 3

The Christians

- *The Christians* (1988): Born Again; Ideal World; One in a Million; Sad Songs; When the Fingers Point; Why Waltz

NA; NA; Interview 3 (1987-1989);
Contemporary Interview 1; Contemporary Interview 2; Contemporary Interview 3

The Housemartins

- *London 0 Hull 4* (1986): Get Up Off Our Knees; Happy Hour; He Ain't Heavy, He's My Brother; I'll Be Your Shelter (Just Like a Shelter); Lean On Me; Over There; People Get Ready; Sheep

Interview 1 (1980-1983); Interview 2 (1984-1986); NA;
NA; NA; NA

The Human League

- *Dare!* (1981): Do or Die; Don't You Want Me; Love Action (I Believe in Love); Open Your Heart; Seconds; The Sound of the Crowd

Interview 1 (1980-1983); Interview 2 (1984-1986); Interview 3 (1987-1989);
Contemporary Interview 1; Contemporary Interview 2; Contemporary Interview 3

The Jam

- *The Gift* (1982): Just Who Is the 5 O'Clock Hero?; Precious; Running on the Spot; The Gift; The Planner's Dream Goes Wrong; Trans-Global Express

Interview 1 (1980-1983); Interview 2 (1984-1986); NA;
Contemporary Interview 1; Contemporary Interview 2; Contemporary Interview 3

The Police

- *Outlandos d'Amour* (1978): Next To You; Roxanne
- *Reggatta De Blanc* (1979): Contact; No Time This Time; Walking on the Moon
- *Zenyatta Mondatta* (1980): Bombs Away; Canary in a Coalmine; Don't Stand So Close to Me; Driven To Tears; When the World Is Running Down
- *Ghost in the Machine* (1981): Every Little Thing She Does Is Magic; Rehumanize Yourself
- *Synchronicity* (1983): Every Breath You Take; King of Pain; O My God; Synchronicity II; Walking in Your Footsteps; Wrapped Around Your Finger

Interview 1 (1980-1983); Interview 2 (1984-1986); Interview 3 (1987-1989);
Contemporary Interview 1; Contemporary Interview 2; Contemporary Interview 3

The Rolling Stones

- *Emotional Rescue* (1980): All About You; Down in The Hole; Send It to Me; She's So Cold; Summer Romance

Interview 1 (1980-1983); Interview 2 (1984-1986); Interview 3 (1987-1989);
Contemporary Interview 1; Contemporary Interview 2; Contemporary Interview 3

The Smiths

- *The Smiths* (1984): Hand In Glove; Miserable Lie; Reel Around the Fountain; Still Ill; What Difference Does It Make; You've Got Everything Now

Interview 1 (1980-1983); Interview 2 (1984-1986); Interview 3 (1987-1989);
Contemporary Interview 1; Contemporary Interview 2; Contemporary Interview 3

The Stone Roses

- *The Stone Roses* (1989): Bye Bye Badman; Don't Stop; I Am the Resurrection; (Song for My) Sugar Spun Sister; Waterfall

Interview 1 (1980-1983); Interview 2 (1984-1986); Interview 3 (1987-1989);
Contemporary Interview 1; Contemporary Interview 2; Contemporary Interview 3

The Style Council

- *Café Bleu* (1984): A Gospel; Headstart for Happiness; Here's One That Got Away; My Ever-Changing Moods; The Paris Match; The Whole Point Of No Return; You're The Best Thing

Interview 1 (1980-1983); Interview 2 (1984-1986); Interview 3 (1987-1989); Contemporary Interview 1; Contemporary Interview 2; Contemporary Interview 3

Thompson Twins

- *Into the Gap* (1984): Hold Me Now; Sister of Mercy; Who Can Stop the Rain; You Take Me Up

Interview 1 (1980-1983); Interview 2 (1984-1986); Interview 3 (1987-1989); Contemporary Interview 1; Contemporary Interview 2; Contemporary Interview 3

Wham!

- *Fantastic* (1983): Bad Boys; Club Tropicana; Come On!; Love Machine
- *Make it Big* (1984): Careless Whisper; Everything She Wants; Freedom; If You Were There

Interview 1 (1980-1983); Interview 2 (1984-1986); Interview 3 (1987-1989); Contemporary Interview 1; Contemporary Interview 2; NA

A.3 Chapter 7: Acoustic analysis

A.3.1 Albums included in the corpus

Bruce Springsteen

- *The River* (1980)
- *Nebraska* (1982)
- *Born in the USA* (1984)
- *Tunnel of Love* (1987)

Duran Duran

- *Duran Duran* (1981)
- *Rio* (1982)
- *Seven and the Ragged Tiger* (1983)
- *Notorious* (1986)
- *Big Thing* (1988)

Billy Joel

- *Glass Houses* (1980)
- *The Nylon Curtain* (1982)
- *An Innocent Man* (1983)
- *The Bridge* (1986)
- *Storm Front* (1989)

Phil Collins

- *Face Value* (1981)
- *Hello, I Must Be Going!* (1982)
- *No Jacket Required* (1985)
- *...But Seriously* (1989)

Judas Priest

- *British Steel* (1980)
- *Point of Entry* (1981)
- *Screaming for Vengeance* (1982)
- *Defenders of the Faith* (1984)
- *Turbo* (1986)
- *Ram it Down* (1988)

Iron Maiden

- *The Number of the Beast* (1982)
- *Piece of Mind* (1983)
- *Powerslave* (1984)
- *Somewhere in Time* (1986)
- *Seventh Son of a Seventh Son* (1988)

B Production Experiment 1: lip tracking

B.1 Ethics Committee Approval

CER U-Paris Cité

(**C**omité d'**É**thique de la **R**echerche)

Présidente : Jacqueline Fagard (jacqueline.fagard@u-paris.fr)
Vice-présidente : Anne-Caroline Fiévet (anne-caroline.fievet@u-paris.fr)

N° 2023-41-CAILLOL-FERRAGNE

PROTOCOLE : Articulatory Constraints in Singing: An electropalatographic & lip tracking study.

Noms du/des chercheur(s) : Coline Caillol ; Emmanuel Ferragne

Email pour la correspondance : coline.caillol@u-paris.fr; emmanuel.ferragne@u-paris.fr

Labo / service : CLILLAC-ARP: Centre de Linguistique Inter-langues, de Lexicologie, de Linguistique Anglaise et de Corpus-Atelier de Recherche sur la Parole; Université Paris Cité

Évalué à la séance du 02/05/2023

AVIS : Favorable

L'avis du CER U-Paris Cité n'exonère pas des formalités réglementaires. A cet égard, il vous appartient notamment, si vous traitez des données se rapportant à un individu directement ou indirectement identifiable, de vous conformer au règlement européen sur la protection des données (RGPD) en vigueur depuis 2018. Pour cela, vous devez solliciter les conseils du/de la DPD (délégué.e. à la protection des données) de votre établissement.

Les investigateurs faisant appel aux services du CER U-Paris Cité s'engagent à lui signaler tout événement non anticipé survenant en cours d'étude. Ces éléments seront utilisés aux fins d'amélioration des futurs services et conseils que le CER U-Paris Cité pourrait donner.

N° IRB : 00012023-41

Jacqueline Fagard

B.2 lip tracking by color segmentation MATLAB script

Lip tracking by color segmentation

This script can be used to process a video file of (preferably colored) lips and compute the vertical and horizontal dimensions of a bounding box for each token. It requires the Image Processing Toolbox.

Other files needed:

- Lip video
- A .txt file containing the lip video file name, tokens and corresponding time stamps
- Function mask2BoundingBox (defined at the end of the LiveScript)

Initialization

Import data

```
% Upload table
tableBornes = readtable('timeStamps.txt');
verticalLength = zeros(size(tableBornes,1),1);
horizontalLength = zeros(size(tableBornes,1),1);

% Find corresponding video file based on first entry in table
fileName = char(tableBornes.file(1));
videoObject = VideoReader(fileName);
```

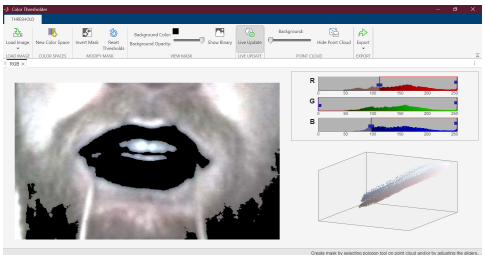
Manually crop a region from a frame

Displays a random frame (here, frame number 1754) from the video, and lets the user crop a region manually from the frame. The region should be big enough to emcompass all lip movement throughout the video, but small enough that there is no superfluous information, especially of a color too similar to that of the lips.

```
[newFrame, rectcrop] = imcrop(read(videoObject,1754));
imshow(newFrame);
% Save image
imwrite(newFrame, 'calibratingFrame.png');
```

Create mask from color space

To do this, upload the calibrating frame from above to the Color Thresholder app of the Image Processing and Computer Vision toolbox. Play around with the color spaces until an appropriate one is found that properly isolates the lip color from the rest of the image and export the resulting mask as a function.



Process video

Loop through all tokens, apply mask to cropped region and create a maximum bounding box from mask.

```
for idx = 1:size(tableBornes,1)
    trueFrameRate = videoObject.NumFrames/videoObject.Duration;
    frame1 = round(trueFrameRate * tableBornes.tmin(idx));
    frame2 = round(trueFrameRate * tableBornes.tmax(idx));
    tempHeight = zeros(frame2-frame1+1, 1);
    tempWidth = zeros(frame2-frame1+1, 1);

    % For all frames between tmin and tmax, crop to the region selected earlier
    % and apply the mask (createMask) exported from the Color Thresholder app
    for jdx = frame1:frame2
        currentFrame = read(videoObject, jdx);
        newFrame = imcrop(createMask(currentFrame), rectcrop);
        % Get the bounding box and its dimensions for the masked region
        % (lips)
        [boundingBox, rectout] = mask2BoundingBox(newFrame);
        % Save the bounding box image for each frame, labeled by word and frame
number
        tempHeight(jdx) = rectout(4);
        tempWidth(jdx) = rectout(3);
        wordStr = char(tableBornes.word(idx));
        tempFileName = [fileName(:,1:end-4) '_' wordStr '_'
num2str(tableBornes.tmin(idx)) '_' num2str(jdx) '.png'];
        tempFileName1 = ['C:\Users\Coline
Caillol\Desktop\PhD\XP_EPG\Analysis\s01\Lips\', tempFileName];
        imwrite(boundingBox, tempFileName1);
    end
```

The image output for each frame should look like this:



Extract dimensions

```
    % Find the maximum height and width of the bounding box across the
    % word's frames and store their values

    [maxHeight, maxHeightIdx] = max(tempHeight);
    [maxWidth, maxWidthIdx] = max(tempWidth);

    verticalLength(idx) = maxHeight;
    horizontalLength(idx) = maxWidth;

    maxHeightFrame(idx) = maxHeightIdx;
    maxWidthFrame(idx) = maxWidthIdx;

end
```

Save max dimensions to file

```
% Save maximum values to file
newTable = [tableBornes, table(verticalLength), table(horizontalLength),
table(maxHeightFrame')];
newTable.Properties.VariableNames{'Var1'} = 'maxHeightFrame';
newTable = [newTable, table(maxWidthFrame')];
newTable.Properties.VariableNames{'Var1'} = 'maxWidthFrame';
writetable(newTable, 'lips_s00.txt', 'Delimiter', '\t');
```

Additional function

mask2BoundingBox

```
function [newIm, rectout] =  mask2BoundingBox(binaryImage)
% Invert mask
revMask = imcomplement(binaryImage);
% Fill gaps
filledMask = imfill(revMask,'holes');
% Keep one object only
oneObj = bwareafilt(filledMask,1);
```



```
% Measurements
s = regionprops(oneObj, 'BoundingBox');
% Prepare final image
newIm = insertShape(uint8(oneObj*255), 'Rectangle', s.BoundingBox);
rectout = s.BoundingBox;
```

B.3 lip tracking by facial landmark recognition Python script

Lip Tracking using Face Mesh in Python

This notebook processes videos to track lips using MediaPipe's Face Mesh. It calculates the Euclidean distance between two facial landmarks that correspond to the lips (0 and 17), writes the coordinates and distance to a text file, and saves the annotated video.

Imports

```
import cv2
import mediapipe as mp
import time
import csv
import math
import os
```

FaceMeshDetector Class

This class sets up and runs the MediaPipe Face Mesh model.

```
class FaceMeshDetector():
    def __init__(self, staticMode=False, maxFaces=1, refineLandmarks=False, minDetectionCon
        self.staticMode = staticMode
        self.maxFaces = maxFaces
        self.refineLandmarks = refineLandmarks
        self.minDetectionCon = minDetectionCon
        self.minTrackCon = minTrackCon

        self.mpDraw = mp.solutions.drawing_utils
        self.mpFaceMesh = mp.solutions.face_mesh
        self.faceMesh = self.mpFaceMesh.FaceMesh(self.staticMode, self.maxFaces, self.refin
                                                 self.minDetectionCon, self.minTrackCon)
        self.drawSpec = self.mpDraw.DrawingSpec(thickness=1, circle_radius=1)

    def findFaceMesh(self, img, draw=True):
        self.imgRGB = cv2.cvtColor(img, cv2.COLOR_BGR2RGB)
        self.results = self.faceMesh.process(self.imgRGB)
        facePoints = []
        if self.results.multi_face_landmarks:
            for faceLms in self.results.multi_face_landmarks:
                if draw:
                    self.mpDraw.draw_landmarks(img, faceLms, self.mpFaceMesh.FACEMESH_CONT
                                               self.drawSpec, self.drawSpec)
                for id, lm in enumerate(faceLms.landmark):
                    ih, iw, ic = img.shape
                    x, y = int(lm.x * iw), int(lm.y * ih)
                    facePoints.append([id, x, y])
        return img, facePoints
```

Euclidean Distance Calculation

```
def euclidean_distance(point1, point2):
    return math.sqrt((point1[1] - point2[1]) ** 2 + (point1[2] - point2[2]) ** 2)
```

Process Videos in Folder

This function loops through `.mp4` files in a folder, processes them frame-by-frame, and writes timestamped lip coordinates and distances to text files. It also outputs annotated videos.

```
def process_videos_in_folder(folder_path):
    video_files = [f for f in os.listdir(folder_path) if f.endswith('.mp4')]

    for video_file in video_files:
        video_path = os.path.join(folder_path, video_file)
        output_file = os.path.join(folder_path, f"{os.path.splitext(video_file)[0]}_lipPoi
        output_video = os.path.join(folder_path, f"{os.path.splitext(video_file)[0]}_faceMe

        cap = cv2.VideoCapture(video_path)
        total_frames = int(cap.get(cv2.CAP_PROP_FRAME_COUNT))
        fps = cap.get(cv2.CAP_PROP_FPS)
        frame_duration = 1 / fps  # Duration of each frame in seconds
        print(f'Total number of frames in {video_file}: {total_frames}')

        pTime = 0
        detector = FaceMeshDetector()

        frame_width = int(cap.get(cv2.CAP_PROP_FRAME_WIDTH))
        frame_height = int(cap.get(cv2.CAP_PROP_FRAME_HEIGHT))

        out = cv2.VideoWriter(output_video, cv2.VideoWriter_fourcc(*'XVID'), fps, (frame_w

        with open(output_file, 'w', newline='') as file:
            writer = csv.writer(file, delimiter='\t')
            writer.writerow(['Timestamp (s)', '0coords (x,y)', '17coords (x,y)', 'Euclidean

            frame_number = 0

            while True:
                success, img = cap.read()
                if not success:
                    break

                img, facePoints = detector.findFaceMesh(img)

                cTime = time.time()
                fps_display = 1 / (cTime - pTime)
                pTime = cTime
                cv2.putText(img, f'FPS: {int(fps_display)}', (20, 70), cv2.FONT_HERSHEY_PLA
                cv2.imshow("Image", img)
                out.write(img)
                cv2.waitKey(1)

                points_0 = [point for point in facePoints if point[0] == 0]
                points_17 = [point for point in facePoints if point[0] == 17]

                if points_0 and points_17:
                    point_0 = points_0[0]
                    point_17 = points_17[0]

                    distance = euclidean_distance(point_0, point_17)

                    # Calculate timestamp
                    timestamp = frame_number * frame_duration

                    row = [f"{timestamp:.3f}",  # Format timestamp to 3 decimal places
                           f"{point_0[1]},{point_0[2]}",
                           f"{point_17[1]},{point_17[2]}",
                           distance]
                    writer.writerow(row)

                frame_number += 1

        cap.release()
        out.release()
```

Run the Script

```
if __name__ == "__main__":
    folder_path = "videos/s04_sentences"  # Replace with your actual folder path
    process_videos_in_folder(folder_path)
```

B.4 Participant information form

Université Paris Cité

CLILLAC

PARTICIPANT QUESTIONNAIRE

ID Code [to be completed by the experimenter]:

Why do I need to provide all this information?

The information you provide in this questionnaire is needed to ensure our experiment will be a good fit for you!

Demographic information is needed to make sure all our participants have similar-enough backgrounds so that we may draw more general conclusions from the data we gather in this experiment. Since this is a linguistics study, we need to know a little more about your language history and practices. Because we are specifically focusing on singing, we also need to know more about your level and experience as a singer.

Contact info & Demographics

Name:

Age: 18-24 25-34 35-49 50-64 65+

Sex: Male Female

Email address:

Current job:

Speaking

1. Where do you currently live?

France England Ireland/UK (apart from England) Other

2. In what country were you raised?

France England Ireland/UK (apart from England) Other

3. Where are your parents from?

France England Ireland/UK (apart from England) Other

4. What language(s) do you speak?

French English Other(s)

5. How would you qualify your accent in English? (Scottish, Scouse, Cockney…)

6. Do you have any known vision, hearing, speech or neurological impairments?

   Yes
   No

Singing

Check all that apply:

I make (most of) my living as a singer.
I've had formal training as a singer (music school, private lessons…) for at least three years.
I don't have formal training but have been singing regularly for at least three years.
I perform as a singer regularly (at least once or twice a month).
I am a classical/opera singer.
I sing (mostly) alone.
I sing (mostly) in a band.
I sing (mostly) in a choir.
I sing (mostly) in English.

What type/genre of singing do you practice the most? (Choir, Jazz, Rock…)

Do you know your vocal range (excluding falsetto) and/or what octave you are most comfortable singing at? If so, please indicate below.

Any other information you would like us to know?

B.5 Recruitment flyer

Université Paris Cité

SINGERS WANTED

FOR A LINGUISTICS STUDY ON **THE SINGING VOICE**

If you are:

- A man
- British
- Living in Paris
- A singer

We are looking for you to participate in an experiment on singing voice pronunciation!

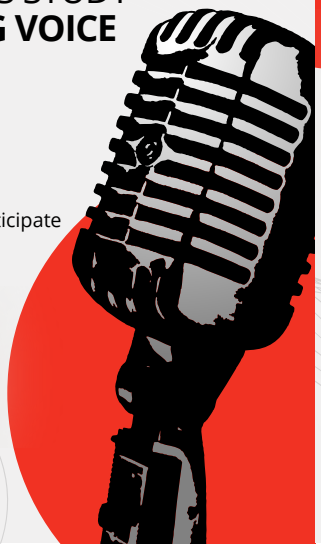
For more information, flip this flyer over or contact:

Coline Caillol

coline.caillol@etu.u-paris.fr

06 38 56 89 52

You can scan here to email directly:



CLILLAC-ARP

Université Paris Cité

What is this for?

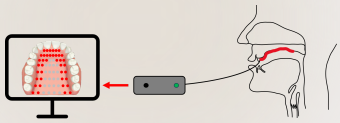
This study is conducted by the linguistics research lab CLILLAC-ARP at Université Paris Cité as part of a doctoral **research project** on pronunciation and articulation in the singing voice.

Who can participate?

We are looking for **English male contemporary music singers** who live in or around Paris and without any known vision, hearing, speech or neurological impairments.

What will I have to do?

Well, sing! But with a catch – we will be using a system called **electropalatography** to record tongue/palate contact. This is what it looks like:



First, we'll take an impression of the roof of your mouth and make your very own custom EPG palate. Then, a few weeks later, you will be asked to come into our lab to sing and read a few sentences for us.

That sounds like a lot! Will I be compensated for my time?

We've tried it ourselves and we promise it's not that bad! But you will get a **100€** gift vouchers if you complete the experiment.

Interested? Want more info?

Don't hesitate to contact us!

Coline Caillol

coline.caillol@etu.u-paris.fr

06 38 56 89 52

B.6 Participant information letter

PARTICIPANT INFORMATION LETTER

Articulatory Constraints in Singing: An electropalatographic & lip tracking study

What's the purpose of this experiment?

Though we can't tell you exactly what it's about just yet, this experiment is meant to study pronunciation in the singing voice, and what strategies singers use to cope with different constraints linked to singing and articulation.

Can I participate in this experiment?

If you are British (preferably from England), a man, and a contemporary music singer (professional or semi-professional), then the answer is yes!

By professional or semi-professional we mean that **at least one** of these points apply to your situation:

- You make (most of) your living as a singer and/or musician
- You've had formal training (conservatory, music school, singing lessons, etc.) for more than 3 years as a singer
- You sing regularly (live performances, recording sessions, etc.)

The study itself focuses on Pop/Rock genres, but any contemporary music singer is welcome. Unfortunately, we cannot accept classical or opera singers.

In this experiment, you will be asked to listen, read and sing. If you have any known vision, hearing, speech or neurological impairments, then this study is probably not the best fit for you.

What does this experiment entail?

In short, you will have to sing and read some sentences aloud, while connected to an electropalatography system and wearing an adjustable ultra-light headset to which a micro-camera is attached. You will also be wearing lipstick. See pictures below.

> *The electropalatography system, or EPG, is a technique used to measure tongue contact against the hard palate during articulation. This is a non-invasive and risk-free technique which requires an impression of the upper teeth and palate to be taken (using a harmless intra-oral scanner), in order to create your own custom artificial palate, fitted with electrodes. You will be asked to put the artificial palate in your mouth and speak and sing. The artificial palate will not impair your singing or speaking abilities. This technique does not present any dangers or known objections.*
>
> *The headset is meant to point the camera at your lips and keep it stable despite movements. It will be used to track your lips. The green lipstick is meant to facilitate automatic lip-tracking.*

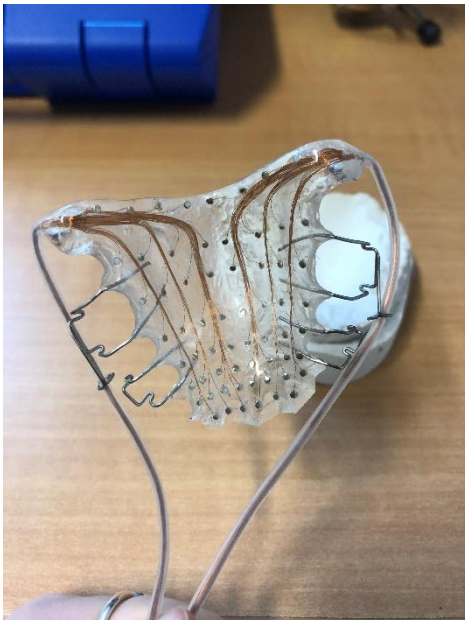
There are three steps to the experiment:

1. The first session will take place at Laboratoire Fraisse (47 rue Alexandre Dumas, 75011), where a licensed dental prosthetist will take an impression of your palate and upper teeth. It is a quick, painless, and totally harmless procedure. This should take 20-30 minutes.

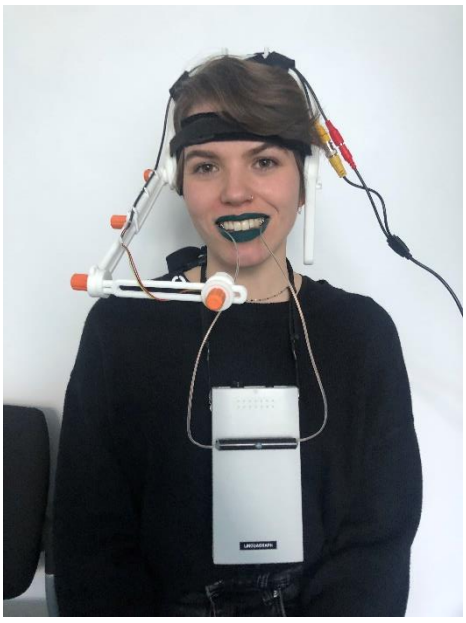
2 Once the impression of your palate has been taken, the next step is to prepare yourself for the day of the experiment. You will need to identify and communicate to the experimenter the highest note you can sing while still being somewhat comfortable. As such, it shouldn't be the highest note you can reach overall, but the highest one you can reach without too much straining. You can use an online pitch detector (https://www.onlinemictest.com/tuners/pitch-detector/) or virtual piano (https://piano.starrynets.com/) if needed. Based on this, we will determine and send you a total of six 10-second melodies to familiarize yourself with what you will have to sing on experiment day: three at a designated high pitch, and the same three an octave lower.

3 You will be contacted for the third and final step about 4 weeks after the first, once your EPG palate is ready. This session will take place at the Olympe de Gouges building of Université Paris Cité (8 place Paul Ricoeur, 75013), in room 318. This is when the actual experiment will occur. You will be standing in a sound-proof recording booth, with the artificial palate in your mouth and wearing the main component of the EPG system around your neck, as well as the camera-equipped headset. You will also be wearing green lipstick to facilitate lip-tracking. You will hear a short instrumental track and melody and you will be shown some lyrics, then when you are ready you will be asked to sing the lyrics based on the melody you've heard. You will also be asked to read different sentences. If you're interested, there will be a debrief with the experimenter explaining the aim of the study. This session will take an hour and a half to two hours.

EPG custom-made palate



Experiment equipment



What are the risks?

There are no physical or psychological risks associated with participation in this experiment.

The EPG system that will be used is non-invasive and safe, though a slight discomfort due to having a foreign object in the mouth may be felt by the participant in the first 10 minutes of wearing the palate. The headset is fully adjustable and very light so that it will fit comfortably on your head. The lipstick

will be patch-tested on the arm to check for any allergic reaction, and applied with a single-use, disposable applicator.

What are the benefits?

There are no personal benefits to your participation in this experiment, though we hope to learn and understand more about singing-voice pronunciation.

If you complete the experiment in its entirety, you will receive compensation in the form of a 100€ gift voucher (can be used in many different places such as Carrefour or Fnac[1]) to thank you for your time and participation.

Will my identity be known or shared?

All the information you share about yourself will be kept confidential.

Information and data collected in this study may be used in publications (written or oral), for other research studies, or shared with other researchers for future research. If this happens, information that could identify participants will be removed before anything is shared.

It is important to note, however, that the recordings collected during the experiment contain your voice and videos of your lips which can be enough to lead to your identification even if your name will never be associated to it explicitly.

What will happen to the results of this experiment?

Data collection, storage, protection, and destruction are conducted in accordance with French and European laws on privacy. All data provided (including consent forms and questionnaires) will be securely saved on a designated external hard drive and stored in a secure locker at the CLILLAC-ARP lab. Participants' names will not appear on any file related to this study. The information participants share will be kept confidential by assigning encrypted ID codes.

Research data will be securely analyzed and preserved from malevolent use. Only the research team will have access to study data. Research data will be retained for 10 years after the end of the project. If participants do not complete the experiment, their data will be immediately destroyed.

If you would like to have access to the results, we will let you know once the study is complete.

***

All research is voluntary. If you decide to participate, you can change your mind later and decide to leave the experiment at any point. If you decide to withdraw, you will not have to give a reason why, nor will you be penalized for it. If you complete the experiment, you can decide to withdraw from the study and ask to have your data deleted within four weeks after your participation.

***

Nevertheless, the custom-made equipment we use is costly, so we do ask that you **take the time to think over the potential constraints of the study** and decide to **participate only if you are genuinely willing to complete the study**.

[1] You can find the full list of eligible stores here: https://boutiques.cheque-cadhoc.fr/enseignes

***

Please let the experimenter know by email if you're interested in participating, so we can start to schedule the sessions!

In short

If you fit the profile and agree to participate in the study:

1st session [30 min]: Palate impression taken at Laboratoire Fraisse, 47 rue Alexandre Dumas 75011.

Between the two sessions, practice singing three 10 second melodies at your own pace.

2nd session [2h]: About a month later, experiment at Université Paris Cité, Olympe de Gouges building, 8 Place Paul Ricoeur. You will be asked to sing and speak, while monitored by different instruments.

There are no risks to this study, and no personal benefits though compensation of 100€ worth of gift vouchers will be offered upon completion. All private information about yourself will be kept confidential. The results will be anonymous and stored securely.

CONTACT INFORMATION

Main investigator:

Coline CAILLOL, PhD candidate at Université Paris Cité (CLILLAC-ARP)

coline.caillol@etu.u-paris.fr ; coline.caillol@gmail.com

O6.38.56.89.52

Supervisor:

Emmanuel FERRAGNE, Professor at Université Paris Cité (CLILLAC-ARP)

emmanuel.ferragne@u-paris.fr

Laboratoire Fraisse

47, rue Alexandre Dumas 75011 PARIS

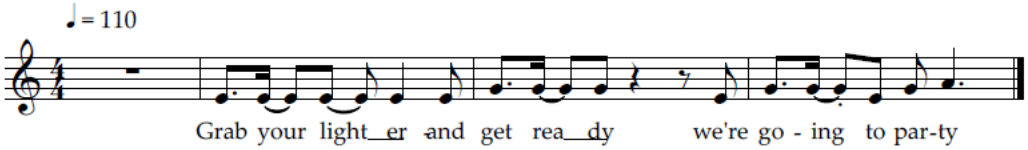
Centre de Linguistique Inter-langues, de Lexicologie, de Linguistique Anglaise et de Corpus-Atelier de Recherche sur la Parole (CLILLAC-ARP, EA3967).

8, place Paul Ricoeur 75013 PARIS

B.7 Melody scores

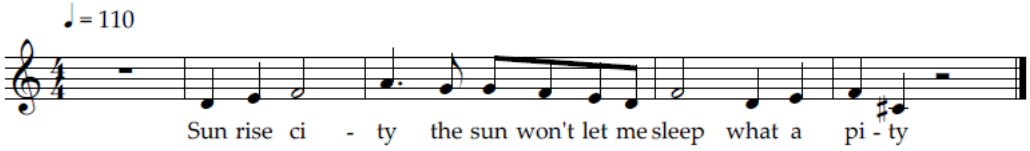
Melody 1

grabLighter



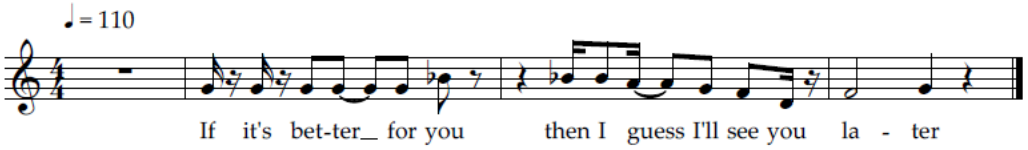
Melody 2

sunriseCity



Melody 3

ifBetter



B.8 Experiment stimuli display MATLAB script

Lip-tracking and EPG experiments display

This script requires Psychtoolbox. In part 1, based on given pitch parameters, it retrieves, randomizes and displays two versions of karaoke videos (one with a synthetic voice singing the melody, and an instrumental-only version). It also records the participant during the instrumental-only video and provides both live feedback and the opportunity to listen to the recording right after. The participant can navigate through the videos using a clicker. In part 2, retrieves, randomizes, and displays the previously-sung lyrics on the screen to be read. It provides instruction to the participant throughout (see screenshots of the different outputs) and writes all the experiment metadata to a dedicated file.

Other files needed:

- Karaoke videos (melodic and instrumental versions)
- List of stimuli

Pre-experiment initialization

Locating, preparing and randomizing video files

```
%Open up dialog box to enter regular & high pitch values
prompt = {'Trial name'};
prompt1 = {'Regular pitch:'};
prompt2 = {'High pitch:'};
trialTitle = 'Trial name';
dlgtitle = 'Pitch Selection';

trialNameCell = inputdlg(prompt, trialTitle, [1 50]);
trialName = char(trialNameCell);
regPitch = inputdlg(prompt1, dlgtitle, [1 50]);
regPitchName = char(regPitch);
highPitch = inputdlg(prompt2, dlgtitle, [1 50]);
highPitchName = char(highPitch);

% Display the user's input for both fields
disp(['User entered for Field 1: ', regPitch]);
disp(['User entered for Field 2: ', highPitch]);

% Instrumentals folder
instrumentalFiles = 'C:\Users\Coline
Caillol\Desktop\PhD\XP_EPG\Stimuli\PsychToolBox\karaokeVidsNew\Instrumentals';

% Set up myFolder
myFolder = 'C:\Users\Coline
Caillol\Desktop\PhD\XP_EPG\Stimuli\PsychToolBox\karaokeVids';

% Fetch list of all files in corresponding subfolder of value given to regPitch
prompt
% Ex: if regular pitch = D2, fetches all files in the D2 folder
```



```
regPitchFolder = fullfile(myFolder, regPitch);
regPitchFolder = char(regPitchFolder);
regPitchFiles = dir(fullfile(regPitchFolder, '*.mp4'));

disp("Files for regular pitch:")
for i = 1:length(regPitchFiles)
    disp(regPitchFiles(i).name);
end

% Fetch list of all files in corresponding subfolder of value given to highPitch
prompt
highPitchFolder = fullfile(myFolder, highPitch);
highPitchFolder = char(highPitchFolder);
highPitchFiles = dir(fullfile(highPitchFolder, '*.mp4'));

disp("Files for high pitch:")
for i = 1:length(highPitchFiles)
    disp(highPitchFiles(i).name);
end

% Combine reg pitch and high pitch files into a new temporary folder called
% tempFolder
tempFolder = tempname;
mkdir(tempFolder);

for i = 1:length(regPitchFiles)
    sourceFile = fullfile(regPitchFolder, regPitchFiles(i).name);
    destinationFile = fullfile(tempFolder, regPitchFiles(i).name);
    copyfile(sourceFile, destinationFile);
end

for i = 1:length(highPitchFiles)
    sourceFile = fullfile(highPitchFolder, highPitchFiles(i).name);
    destinationFile = fullfile(tempFolder, highPitchFiles(i).name);
    copyfile(sourceFile, destinationFile);
end

disp(['Files copied successfully to temporary folder: ' tempFolder]);
disp('Contents of the temporary folder:');
tempFolderContents = dir(fullfile(tempFolder, '*.mp4*'));
for i = 1:length(tempFolderContents)
    disp(tempFolderContents(i).name);
end

% Randomize file order in temporary folder
shuffledKaraoke = randperm(length(tempFolderContents));

% Display the shuffled order of video files
disp('Randomized Order of Video Files:');
disp({tempFolderContents(shuffledKaraoke).name});
```


Write metada and save workspace

```
% Write meta data to .txt file (regular pitch, high pitch, randomized stim
% list)
stimOrderPath = [trialName '_metaData.txt'];
fid = fopen(stimOrderPath, 'w');
fprintf(fid, 'Regular pitch: ');
regularPitch = regPitchName;
fprintf(fid, '%s\n', regularPitch);
fprintf(fid, 'High pitch: ');
highPitch = highPitchName;
fprintf(fid, '%s\n', highPitch);
fprintf(fid, 'Order of stimuli:\n');
for i = 1:length(shuffledKaraoke)
    fileIndex = shuffledKaraoke(i);
    fileIdx = num2str(i);
    fileName = tempFolderContents(fileIndex).name;
    fileAll = [fileIdx '-' fileName];
    fprintf(fid, '%s\n', fileAll);
end
fclose(fid);

workspaceName = [trialName 'Workspace.mat'];
save(workspaceName);
```

Setting up Pyschtoolbox

```
rrect=[500 420 1420 980];

% Call default settings for Psychtoolbox
KbName('UnifyKeyNames');
Screen('Preference','VisualDebugLevel', 3);
Screen('Preference', 'SkipSyncTests', 1);

% Get the screen numbers
screens = Screen('Screens');
screenNumber = max(screens);

% Define colors
black = BlackIndex(screenNumber);
white = WhiteIndex(screenNumber);

% Initialize sound
InitializePsychSound;

% Open an on screen window
[window, windowRect] = Screen('OpenWindow', screenNumber, white);
```



```
% Get the vertical refresh rate of the monitor
ifi = Screen('GetFlipInterval', window);
% Set the blend funciton for the screen
Screen('BlendFunction', window, 'GL_SRC_ALPHA', 'GL_ONE_MINUS_SRC_ALPHA');
% Get the size of the on screen window in pixels
[screenXpixels, screenYpixels] = Screen('WindowSize', window);
% Get the centre coordinate of the window in pixels
[xCenter, yCenter] = RectCenter(windowRect);
% Maximum priority level
topPriorityLevel = MaxPriority(window);
Priority(topPriorityLevel);
% Flip to the vertical retrace rate
vbl = Screen('Flip', window);

% Set text properties
Screen('TextFont', window, 'Calibri');
Screen('TextColor', window, [0 0 0]);
textEncoding = Screen('Preference', 'TextEncodingLocale', 'UTF-8');
```

Setting up other devices

```
% Set-up the audio recorder for post-recording feeback
recorder = audiorecorder(44100, 16, 1);

% Find and open the clicker to go through slides
for i = 1:length(devices)
    if strcmp(devices(i).product, 'USB Receiver')
        gamepadIndex = i;
        vendorID = devices(i).vendorID;
        productID = devices(i).productID;
        break;
    end
end
if isempty(gamepadIndex)
    error('Gamepad not found');
end
```

Key mapping for Logitech clicker:

- > = PageDown
- < = PageUp
- Right button = \?

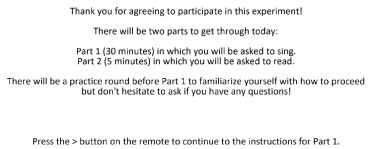
Practice round

Instructions and set-up

```
% First message
 firstmessage = ['\n Thank you for agreeing to participate in this experiment!' ...
```

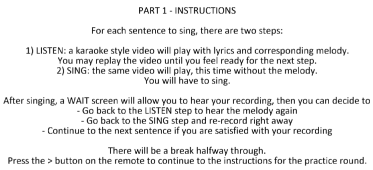
```
    '\n' ...
    '\n There will be two parts to get through today: ' ...
    '\n' ...
    '\n Part 1 (30 minutes) in which you will be asked to sing.' ...
    '\n Part 2 (5 minutes) in which you will be asked to read.' ...
    '\n' ...
    '\n There will be a practice round before Part 1 to familiarize yourself with
how to proceed' ...
    '\n but don''t hesitate to ask if you have any questions!' ...
    '\n' ...
    '\n' ...
    '\n' ...
    '\n' ...
    '\n Press the > button on the remote to continue to the instructions for Part
1.'];

Screen('TextSize', window, 60);
DrawFormattedText(window, firstmessage, 'center', 'center');
Screen('Flip', window);

while true
    [~, ~, keyCode] = KbCheck;
    if keyCode(KbName('PageDown'))
        % Wait until the key is released
        while KbCheck
            % Do nothing, just wait
        end
        WaitSecs(0.1);
        break
    end
end
```

Output:




```
% Display preliminary instructions & wait for mouse click
message = [
    '\n PART 1 - INSTRUCTIONS' ...
    '\n' ...
    '\n For each sentence to sing, there are two steps: ' ...
    '\n' ...
    '\n 1) LISTEN: a karaoke style video will play with lyrics and corresponding
melody.' ...
    '\n You may replay the video until you feel ready for the next step.' ...
    '\n 2) SING: the same video will play, this time without the melody. ' ...
    '\n You will have to sing.' ...
    '\n ' ...
    '\n After singing, a WAIT screen will allow you to hear your recording, then
you can decide to ' ...
    '\n - Go back to the LISTEN step to hear the melody again' ...
    '\n - Go back to the SING step and re-record right away' ...
    '\n - Continue to the next sentence if you are satisfied with your recording'
...
    '\n' ...
    '\n There will be a break halfway through. '...
    '\n Press the > button on the remote to continue to the instructions for the
practice round.'];
Screen('TextSize', window, 60);
DrawFormattedText(window, message, 'center', 'center');
Screen('Flip', window);

% Wait for right arrow (button 1 on gamepad) key *release*
while true
    [~, ~, keyCode] = KbCheck;
    if keyCode(KbName('PageDown'))
        % Wait until the key is released
        while KbCheck
            % Do nothing, just wait
        end
        WaitSecs(0.1);
        break
    end
end
```

Output:




```
%% Practice instructions

messagePrac = ['\n PRACTICE ROUND - INSTRUCTIONS' ...
    '\n' ...
    '\n During this round, you will do steps LISTEN and SING twice,' ...
    '\n once for each selected pitch (normal and high).' ...
    '\n' ...
    '\n You will be able to confirm or modify the pitch choice.' ...
    '\n ' ...
    '\n You may also adjust the microphone settings in front of you' ...
    '\n for overall volume and mix control.'...
    '\n'...
    '\n Press > to start the practice round'];

messageC = 'Press > to start the practice round.';

imagePath = 'C:\Users\Coline
Caillol\Desktop\PhD\XP_EPG\Stimuli\PsychToolBox\micControls1.png';
image = imread(imagePath);

% Get the size of the resized image
[imageWidth, imageHeight] = size(image);

newImageWidth = 800;
newImageHeight = 921;

[ScreenWidth, ScreenHeight] = Screen('WindowSize', window);

% Convert the resized image to a texture
texture = Screen('MakeTexture', window, image);

% Get screen dimensions
rect = Screen('Rect', window);
```



```
% Define normalized position of image on screen
imageRect = [0.36, 0.35, 0.36 + newImageWidth / rect(3), 0.35 + newImageHeight /
rect(4)];
imageRectnorm = imageRect .* [rect(3), rect(4), rect(3), rect(4)];

% Display on screen
Screen('TextSize', window, 50);
DrawFormattedText(window, messagePrac, 'center', 50);
%DrawFormattedText(window, messageC, 'center', 1585);
Screen('DrawTexture', window, texture, [], imageRectnorm);

Screen('Flip', window);
```

Ouput:



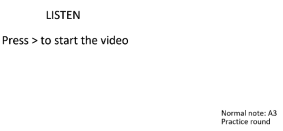

```
% Wait for right arrow (button 1 on gamepad) key *release*
while true
    [~, ~, keyCode] = KbCheck;
    if keyCode(KbName('PageDown'))
        % Wait until the key is released
        while KbCheck
            % Do nothing, just wait
        end
        WaitSecs(0.1);
        break
    end
end
```

Practice round trials loop

Practice loop - Play one random regular pitch and one random high pitch video then ask if chosen pitch is okay. If yes, move on to end of practice round screen.

```
% Display LISTEN screen with selected pitch indicated in the bottom corner,
% wait for a click before continuing
message = ['LISTEN ' ...
    '\n'...
    '\n Press > to start the video'];
Screen('TextSize', window, 75);
DrawFormattedText(window, message, 'center', 'center');
practiceMess = 'Practice round';
pitchMess = ['Normal note: ' regPitchName];

rect = Screen('Rect', window);
xPos = rect(3) - 500;
yPos = rect(4) - 250;

Screen('TextSize', window, 50);
DrawFormattedText(window, practiceMess, xPos, yPos);
DrawFormattedText(window, pitchMess, xPos, yPos - 50);
Screen('Flip', window);
```

Output:




```
% Wait for right arrow (button 1 on gamepad) key *release*
while true
    [~, ~, keyCode] = KbCheck;
    if keyCode(KbName('PageDown'))
        % Wait until the key is released
        while KbCheck
            % Do nothing, just wait
        end
```



```
        WaitSecs(0.1);
        break
    end
end

% Access regular pitch random video

% Randomize contents of regular pitch videos folder only
shuffledRegular = randperm(length(regPitchFiles));

for j = 1:1

    practiceVideoReg = regPitchFiles(shuffledRegular(j));
    practiceVideoRegPath = fullfile(regPitchFolder, practiceVideoReg.name);

    % Get corresponding instrumental
    baseName = regexprep(practiceVideoReg.name, '_.*$', '');
    pitchName = regexprep(practiceVideoReg.name, '.*_(.*?)\..*', '$1');
    practiceInstrumentalVideoRegName = [pitchName baseName '_instrumental.mp4'];
    practiceInstrumentalVideoRegPath = fullfile(instrumentalFiles,
practiceInstrumentalVideoRegName);

    % Get audio using ffmpeg
    practiceRegAudio = 'tempPracticeRegAudio.wav';

    ffmpegCommand = sprintf('ffmpeg -i "%s" -vn -acodec pcm_s16le -ar 44100 -ac 2
"%s"', practiceVideoRegPath, practiceRegAudio);
    % Execute the ffmpeg command in the background using the ! operator
    [~, ~] = system(['start /B cmd /C "', ffmpegCommand, '" >NUL 2>&1']);

    % Open video
    practiceRegKaraoke = Screen('OpenMovie', window, practiceVideoRegPath);

    % Set up audio using Psychtoolbox
    InitializePsychSound;
    [z, ~] = audioread(practiceRegAudio);
    audioHandleRegPractice = PsychPortAudio('Open', [], [], 0, [], 2);

    % Start video playback
    Screen('PlayMovie', practiceRegKaraoke, 1);

    % Start audio playback
    PsychPortAudio('FillBuffer', audioHandleRegPractice, z');
    PsychPortAudio('Start', audioHandleRegPractice, 1, 0, 1);

    % Display frames until the end of the video
    while true
        [~, ~, keyCode] = KbCheck;
       % Check if a specific key is pressed (e.g., 'ESCAPE' key)
        if keyCode(KbName('space'))
```

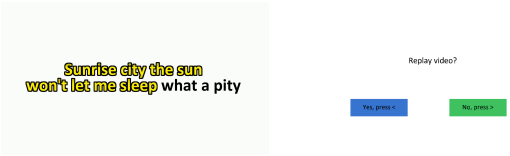


```
            % Exit the loop if 'space' key is pressed
            break
        end
        tex = Screen('GetMovieImage', window, practiceRegKaraoke);
        if tex > 0
            % Draw the current frame
            Screen('DrawTexture', window, tex);
            Screen('Flip', window);
            % Close the texture to release resources
            Screen('Close', tex);
        else
            % End of the video
            break;
        end
    end

    % Stop video and audio playback and close ressources
    Screen('PlayMovie', practiceRegKaraoke, 0);
    PsychPortAudio('Stop', audioHandleRegPractice);

    % Display the replay question & buttons
    messageRep = 'Replay video?';
    buttonRep1Text = 'Yes, press <';
    buttonRep2Text = 'No, press >';

    buttonRepWidth= 500;
    buttonRepHeight = 150;

    % Get the screen dimensions
    screenRect = Screen('Rect', window);

    % Define normalized coordinates for the buttons
    buttonRep1Rect = [0.25, 0.7, 0.25 + buttonRepWidth / screenRect(3), 0.7 +
buttonRepHeight / screenRect(4)];
    buttonRep2Rect = [0.55, 0.7, 0.55 + buttonRepWidth / screenRect(3), 0.7 +
buttonRepHeight / screenRect(4)];

    buttonRep1Rectnorm = buttonRep1Rect .* [screenRect(3), screenRect(4),
screenRect(3), screenRect(4)];
    buttonRep2Rectnorm = buttonRep2Rect .* [screenRect(3), screenRect(4),
screenRect(3), screenRect(4)];

    % Draw buttons
    Screen('FillRect', window, [55, 116, 207], buttonRep1Rectnorm);
    Screen('FillRect', window, [62, 193, 94], buttonRep2Rectnorm);

    Screen('TextSize', window, 75);
    DrawFormattedText(window, messageRep, 'center', 'center');

    Screen('TextSize', window, 60);
```



```
    DrawFormattedText(window, buttonRep1Text, 'center', 'center', [0 0 0], [], [],
[], [], [], buttonRep1Rectnorm);
    DrawFormattedText(window, buttonRep2Text, 'center', 'center', [0 0 0], [], [],
[], [], [], buttonRep2Rectnorm);

    Screen('Flip', window);
```

Output: Instrumental video (left), Replay screen (right)



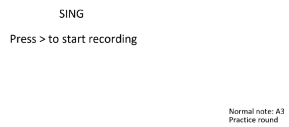

```
    while true

        % If < is pressed on clicker, go back
        [~, ~, keyCode] = KbCheck;
        if keyCode(KbName('PageUp'))
            % Wait until the key is released
            while KbCheck
                % Do nothing, just wait
            end
            WaitSecs(0.1);
            % Set movie time to the beginning and replay the video
            Screen('SetMovieTimeIndex', practiceRegKaraoke, 0);
            Screen('PlayMovie', practiceRegKaraoke, 1);
            PsychPortAudio('FillBuffer', audioHandleRegPractice, z');
            PsychPortAudio('Start', audioHandleRegPractice, 1, 0, 1);
            % Display frames until the end of the video
            while true
                [~, ~, keyCode] = KbCheck;
                % Check if a specific key is pressed (e.g., 'ESCAPE' key)
                if keyCode(KbName('space'))
                    % Exit the loop if 'space' key is pressed
                    break
                end
                tex = Screen('GetMovieImage', window, practiceRegKaraoke);
                if tex > 0
                    % Draw the current frame
                    Screen('DrawTexture', window, tex);
```

```
                    Screen('Flip', window);
                    % Close the texture to release resources
                    Screen('Close', tex);
                else
                    % End of the video
                    break
                end
            end
            Screen('PlayMovie', practiceRegKaraoke, 0);
            PsychPortAudio('Stop', audioHandleRegPractice);

            Screen('FillRect', window, [55, 116, 207], buttonRep1Rectnorm);
            Screen('FillRect', window, [62, 193, 94], buttonRep2Rectnorm);

            Screen('TextSize', window, 75);
            DrawFormattedText(window, messageRep, 'center', 'center');

            Screen('TextSize', window, 60);
            DrawFormattedText(window, buttonRep1Text, 'center', 'center', [0 0 0],
[], [], [], [], [], buttonRep1Rectnorm);
            DrawFormattedText(window, buttonRep2Text, 'center', 'center', [0 0 0],
[], [], [], [], [], buttonRep2Rectnorm);
            Screen('Flip', window);
        end
        % Brief pause to release key buffer
        WaitSecs(0.1);

        % If > is pressed on clicker - continue
        [~, ~, keyCode] = KbCheck;
        if keyCode(KbName('PageDown'))
            % Wait until the key is released
            while KbCheck
                % Do nothing, just wait
            end
            WaitSecs(0.1);
            break;
        end
    end

    % Close resources
    delete(practiceRegAudio);

    % Move on to recording screen
    message = ['SING ' ...
        '\n'...
        '\n Press > to start recording'];
    Screen('TextSize', window, 75);
    DrawFormattedText(window, message, 'center', 'center');

    Screen('TextSize', window, 50);
```



```
    DrawFormattedText(window, practiceMess, xPos, yPos);
    DrawFormattedText(window, pitchMess, xPos, yPos - 50);

    Screen('Flip', window);
```

Output:




```
    while true
        [~, ~, keyCode] = KbCheck;
        if keyCode(KbName('PageDown'))
            % Wait until the key is released
            while KbCheck
                % Do nothing, just wait
            end
            WaitSecs(0.1);
            break
        end
    end

    % Separate audio and video using ffmepg
    practiceIns = 'practiceIns.wav';

    ffmpegCommand = sprintf('ffmpeg -i "%s" -vn -acodec pcm_s16le -ar 44100 -ac 2
"%s"', practiceInstrumentalVideoRegPath, practiceIns);
    [~, ~] = system(['start /B cmd /C "', ffmpegCommand, '" >NUL 2>&1']);

    % Load Instrumental video
    practiceRegInstrumental = Screen('OpenMovie', window,
practiceInstrumentalVideoRegPath);

    % Set up audio recording from the microphone
```



```
    record(recorder);

    % Set up audio using Psychtoolbox
    [v, ~] = audioread(practiceIns);
    audioHandlePracticeIns = PsychPortAudio('Open', [], [], 0, 44100, 2);

    % Start video playback
    Screen('PlayMovie', practiceRegInstrumental, 1);

    % Start audio playback & recording
    PsychPortAudio('FillBuffer', audioHandlePracticeIns, v');
    PsychPortAudio('Start', audioHandlePracticeIns, 1, 0, 1);

    while true
        [~, ~, keyCode] = KbCheck;
        % Check if a specific key is pressed (e.g., 'ESCAPE' key)
        if keyCode(KbName('space'))
            % Exit the loop if 'space' key is pressed
            break
        end
        tex = Screen('GetMovieImage', window, practiceRegInstrumental);
        if tex > 0
            % Draw the current frame
            Screen('DrawTexture', window, tex);
            Screen('Flip', window);
            % Close the texture to release resources
            Screen('Close', tex);
        else
            % End of the video
            break;
        end
    end

    % Stop video and audio (both playback & recording)
    Screen('PlayMovie', practiceRegInstrumental, 0);
    PsychPortAudio('Stop', audioHandlePracticeIns);
    stop(recorder);

    % Get the recorded audio data
    audioData = getaudiodata(recorder);

    % Save the audio data to a file
    audiowrite('practiceRecordedAudio.wav', audioData, 44100);
    practiceAudioPB = 'practiceRecordedAudio.wav';

    % Move to recording playback screen
    messagePB = ['WAIT'...
        '\n' ...
        '\n Listen to your recording' ...
        '\n'...
```



```
        '\n Press > to continue'];
    Screen('TextSize', window, 75);
    DrawFormattedText(window, messagePB, 'center', 'center');
    Screen('Flip', window);
```

Output: Karaoke video (left), Playback screen (right)



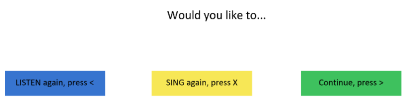

```
    % Initialize recording playback
    [p, fs] = audioread(practiceAudioPB);
    audioHandlePBpractice = PsychPortAudio('Open', [], [], 0, [], 1);

    % Set playback volume
    volumeLevel = 10; % 900% volume increase I think? (1.5% would be 50% volume
increase)
    PsychPortAudio('Volume', audioHandlePBpractice, volumeLevel);

    % Start recording playback & wait for key press to move on
    PsychPortAudio('FillBuffer', audioHandlePBpractice, p');
    PsychPortAudio('Start', audioHandlePBpractice, 1, 0, 1);
```

```
    while true
        [~, ~, keyCode] = KbCheck;
        if keyCode(KbName('PageDown'))
            % Wait until the key is released
            while KbCheck
                % Do nothing, just wait
            end
            WaitSecs(0.1);
            break
        end
    end

    PsychPortAudio('Stop', audioHandlePBpractice);
    PsychPortAudio('Close', audioHandlePBpractice);

    % Record again? screen
    messageR = 'Would you like to... ';
    button1Text = 'LISTEN again, press <';
    button2Text = 'SING again, press X';
    button3Text = 'Continue, press >';

    button1Color = [55, 116, 207];
    button2Color = [247, 231, 86];
    button3Color = [62, 193, 94];

    buttonWidth = 600;
    buttonHeight = 150;

    % Get the screen dimensions
    screenRect = Screen('Rect', window);

    button1Rect = [0.06, 0.7, 0.06 + buttonWidth / screenRect(3), 0.7 +
buttonHeight / screenRect(4)];
    button2Rect = [0.365, 0.7, 0.365 + buttonWidth / screenRect(3), 0.7 +
buttonHeight / screenRect(4)];
    button3Rect = [0.675, 0.7, 0.675 + buttonWidth / screenRect(3), 0.7 +
buttonHeight / screenRect(4)];

    button1Rectnorm = button1Rect .* [screenRect(3), screenRect(4), screenRect(3),
screenRect(4)];
    button2Rectnorm = button2Rect .* [screenRect(3), screenRect(4), screenRect(3),
screenRect(4)];
    button3Rectnorm = button3Rect .* [screenRect(3), screenRect(4), screenRect(3),
screenRect(4)];

    Screen('FillRect', window, button1Color, button1Rectnorm);
    Screen('FillRect', window, button2Color, button2Rectnorm);
    Screen('FillRect', window, button3Color, button3Rectnorm);

    Screen('TextSize', window, 75);
```



```
    DrawFormattedText(window, messageR, 'center', 'center');

    Screen('TextSize', window, 55);
    DrawFormattedText(window, button1Text, 'center', 'center', [0 0 0], [], [], [],
[], [], button1Rectnorm);
    DrawFormattedText(window, button2Text, 'center', 'center', [0 0 0], [], [], [],
[], [], button2Rectnorm);
    DrawFormattedText(window, button3Text, 'center', 'center', [0 0 0], [], [], [],
[], [], button3Rectnorm);
    Screen('Flip', window);
```

Output:




```
    % Wait for button press on clicker
    while true

        % Move on to next stim if > is clicked
        [~, ~, keyCode] = KbCheck;
        if keyCode(KbName('PageDown'))
            % Wait until the key is released
            while KbCheck
                % Do nothing, just wait
            end
            break;
        end
        % Brief pause to release key buffer
        WaitSecs(0.1);

        % Move back to recording screen if little screen is clicked
        [~, ~, keyCode] = KbCheck;
        if keyCode(KbName('/?'))
```



```
            % Wait until the key is released
            while KbCheck
                % Do nothing, just wait
            end
            Screen('SetMovieTimeIndex', practiceRegInstrumental, 0);
            Screen('PlayMovie', practiceRegInstrumental, 1);
            record(recorder);
            PsychPortAudio('FillBuffer', audioHandlePracticeIns, v');
            PsychPortAudio('Start', audioHandlePracticeIns, 1, 0, 1);
            % Display frames until the end of the video
            while true
                [~, ~, keyCode] = KbCheck;
                % Check if a specific key is pressed (e.g., 'ESCAPE' key)
                if keyCode(KbName('space'))
                    % Exit the loop if 'space' key is pressed
                    break
                end
                tex = Screen('GetMovieImage', window, practiceRegInstrumental);
                if tex > 0
                    % Draw the current frame
                    Screen('DrawTexture', window, tex);
                    Screen('Flip', window);
                    % Close the texture to release resources
                    Screen('Close', tex);
                else
                    break
                end
            end
            Screen('PlayMovie', practiceRegInstrumental, 0);
            PsychPortAudio('Stop', audioHandlePracticeIns);
            stop(recorder);

             % Get the recorded audio data
            audioData = getaudiodata(recorder);
            % Save the audio data to a file
            audiowrite('practiceRecordedAudio.wav', audioData, 44100);
            practiceAudioPB = 'practiceRecordedAudio.wav';

            % Move to recording playback screen
            messagePB = ['WAIT'...
                '\n' ...
                '\n Listen to your recording' ...
                '\n'...
                '\n Press > to continue'];
            Screen('TextSize', window, 75);
            DrawFormattedText(window, messagePB, 'center', 'center');
            Screen('Flip', window);

            % Initialize recording playback
            [p, fs] = audioread(practiceAudioPB);
```



```
            audioHandlePBpractice = PsychPortAudio('Open', [], [], 0, [], 1);

            % Set playback volume
            volumeLevel = 10; % 900% volume increase I think? (1.5% would be 50%
volume increase)
            PsychPortAudio('Volume', audioHandlePBpractice, volumeLevel);

            % Start recording playback
            PsychPortAudio('FillBuffer', audioHandlePBpractice, p');
            PsychPortAudio('Start', audioHandlePBpractice, 1, 0, 1);

            while true

                [~, ~, keyCode] = KbCheck;
                if keyCode(KbName('PageDown'))
                    % Wait until the key is released
                    while KbCheck
                        % Do nothing, just wait
                    end
                    WaitSecs(0.1);
                    break
                end
            end

            Screen('PlayMovie', practiceRegInstrumental, 0);
            PsychPortAudio('Stop', audioHandlePracticeIns);
            PsychPortAudio('Stop', audioHandlePBpractice);

            % Go back to selection screen
            Screen('FillRect', window, button1Color, button1Rectnorm);
            Screen('FillRect', window, button2Color, button2Rectnorm);
            Screen('FillRect', window, button3Color, button3Rectnorm);

            Screen('TextSize', window, 75);
            DrawFormattedText(window, messageR, 'center', 'center');

            Screen('TextSize', window, 60);
            DrawFormattedText(window, button1Text, 'center', 'center', [0 0 0], [],
[], [], [], button1Rectnorm);
            DrawFormattedText(window, button2Text, 'center', 'center', [0 0 0], [],
[], [], [], button2Rectnorm);
            DrawFormattedText(window, button3Text, 'center', 'center', [0 0 0], [],
[], [], [], button3Rectnorm);
            Screen('Flip', window);

        end

        % If < is pressed on clicker, go back to listening screen
        [~, ~, keyCode] = KbCheck;
        if keyCode(KbName('PageUp'))
```

```
            % Wait until the key is released
            while KbCheck
                % Do nothing, just wait
            end
            % Set Karaoke movie time to the beginning and replay the video
            Screen('SetMovieTimeIndex', practiceRegKaraoke, 0);
            PsychPortAudio('FillBuffer', audioHandleRegPractice, z');
            PsychPortAudio('Start', audioHandleRegPractice, 1, 0, 1);
            Screen('PlayMovie', practiceRegKaraoke, 1);
            % Display frames until the end of the video
            while true
                [~, ~, keyCode] = KbCheck;
                % Check if a specific key is pressed (e.g., 'ESCAPE' key)
                if keyCode(KbName('space'))
                    % Exit the loop if 'space' key is pressed
                    break
                 end
                tex = Screen('GetMovieImage', window, practiceRegKaraoke);
                if tex > 0
                    % Draw the current frame
                    Screen('DrawTexture', window, tex);
                    Screen('Flip', window);
                    % Close the texture to release resources
                    Screen('Close', tex);
                else
                    % End of the video
                    break
                end
            end
            Screen('PlayMovie', practiceRegKaraoke, 0);
            PsychPortAudio('Stop', audioHandleRegPractice);

            % Go back to selection screen

            Screen('FillRect', window, button1Color, button1Rectnorm);
            Screen('FillRect', window, button2Color, button2Rectnorm);
            Screen('FillRect', window, button3Color, button3Rectnorm);

            Screen('TextSize', window, 75);
            DrawFormattedText(window, messageR, 'center', 'center');

            Screen('TextSize', window, 60);
            DrawFormattedText(window, button1Text, 'center', 'center', [0 0 0], [],
[], [], [], [], button1Rectnorm);
            DrawFormattedText(window, button2Text, 'center', 'center', [0 0 0], [],
[], [], [], [], button2Rectnorm);
            DrawFormattedText(window, button3Text, 'center', 'center', [0 0 0], [],
[], [], [], [], button3Rectnorm);
            Screen('Flip', window);
```



```
        end

        WaitSecs(0.01); % Short delay to avoid high CPU usage
        % end
    end
end

% Close resources
Screen('CloseMovie', practiceRegInstrumental);
Screen('CloseMovie', practiceRegKaraoke)
PsychPortAudio('Close', audioHandlePracticeIns);
delete(practiceRegAudio);
delete(practiceIns);
delete(practiceAudioPB);
```

The exact same portion of code is repeated for the high pitched practice round.

Experimental display

Part 1 - Singing

```
InitializePsychSound;
audioHandleKar = PsychPortAudio('Open', [], [], 0, [], 2);
audioHandleIns = PsychPortAudio('Open', [], [], 0, [], 2);
audioHandlePB = PsychPortAudio('Open', [], [], 0, [], 1);

% % Loop
for i = 1:length(tempFolderContents)
    trialInfo = sprintf('Trial %d of %d ', i, length(tempFolderContents));
    disp(trialInfo);
    trialNum = (i);
```

Checkpoint option

Optional code if the experiment is stopped mid-way through to restart on a specific trial number (do not run if not needed):

```
    % If crash / stopped script --> Run setting up PTB section again, then Run to
End from here.

    trialStopped = 6; % insert the trial number where you want to pick back up

    for i = trialStopped:length(tempFolderContents)
        trialInfo = sprintf('Trial %d of %d ', i, length(tempFolderContents));
        disp(trialInfo);
        trialNum = (i);
```

If not needed, proceed with the rest of the script: similar to the practice round, but loops through all 42 trials.

```
% Display listen wait screen and wait for key press to start playing
    % %video
```



```
    message = ['LISTEN ' ...
        '\n'...
        '\n Press > to continue'];
    Screen('TextSize', window, 75);
    DrawFormattedText(window, message, 'center', 'center');

    rect = Screen('Rect', window);
    xPos = rect(3) - 500; % Set to the right of the screen, adjust as needed
    yPos = rect(4) - 250; % Set to the bottom of the screen, adjust as needed

    Screen('TextSize', window, 50);
    DrawFormattedText(window, trialInfo, xPos, yPos);
    Screen('Flip', window);

    while true
        [~, ~, keyCode] = KbCheck;
        if keyCode(KbName('PageDown'))
            % Wait until the key is released
            while KbCheck
                % Do nothing, just wait
            end
            WaitSecs(0.1);
            break
        end
    end

    % Access the karaoke video using shuffled index
    currentVideo = tempFolderContents(shuffledKaraoke(i));
    % Construct the full path to the video file
    filePath = fullfile(tempFolder, currentVideo.name);

    % Access corresponding instrumental file
    baseName = regexprep(currentVideo.name, '_.*$', '');
    pitchName = regexprep(currentVideo.name, '.*_(.*?)\..*', '$1');
    instrumentalVidName = [pitchName baseName '_instrumental.mp4'];
    instrumentalVidPath = fullfile(instrumentalFiles, instrumentalVidName);

    % Display the current video file name
    disp(['Now Playing: ' currentVideo.name]);

    % Separate audio and video using ffmepg
    audioKar = 'tempAudioKar.wav';
    ffmpegCommand = sprintf('ffmpeg -i "%s" -vn -acodec pcm_s16le -ar 44100 -ac 2
"%s"', filePath, audioKar);
    [~, ~] = system(['start /B cmd /C "', ffmpegCommand, '" >NUL 2>&1']);

    %Load karaoke video
    currentKaraoke = Screen('OpenMovie', window, filePath);
```



```
    % Set up audio using Psychtoolbox
    [z, ~] = audioread(audioKar);

    % Start video playback
    Screen('PlayMovie', currentKaraoke, 1);

    % Start audio playback
    PsychPortAudio('FillBuffer', audioHandleKar, z');
    PsychPortAudio('Start', audioHandleKar, 1, 0, 1);

    % Display frames until the end of the video
    while true
        [~, ~, keyCode] = KbCheck;
        % Check if a specific key is pressed (e.g., 'ESCAPE' key)
        if keyCode(KbName('space'))
            % Exit the loop if 'space' key is pressed
            break
        end
        tex = Screen('GetMovieImage', window, currentKaraoke);
        if tex > 0
            % Draw the current frame
            Screen('DrawTexture', window, tex);
            Screen('Flip', window);
            % Close the texture to release resources
            Screen('Close', tex);
        else
            % End of the video
            break;
        end
    end

    % Stop video and audio playback and close ressources
    Screen('PlayMovie', currentKaraoke, 0);
    PsychPortAudio('Stop', audioHandleKar);

    % Replay option: displays the option to replay the karaoke video

    % Display the replay question & buttons
    messageRep = 'Replay video?';
    buttonRep1Text = 'Yes, press <';
    buttonRep2Text = 'No, press >';

    buttonRepWidth= 500;
    buttonRepHeight = 150;

    % Get the screen dimensions
    screenRect = Screen('Rect', window);

    % Define normalized coordinates for the buttons
```

```
    buttonRep1Rect = [0.25, 0.7, 0.25 + buttonRepWidth / screenRect(3), 0.7 +
buttonRepHeight / screenRect(4)];
    buttonRep2Rect = [0.55, 0.7, 0.55 + buttonRepWidth / screenRect(3), 0.7 +
buttonRepHeight / screenRect(4)];

    buttonRep1Rectnorm = buttonRep1Rect .* [screenRect(3), screenRect(4),
screenRect(3), screenRect(4)];
    buttonRep2Rectnorm = buttonRep2Rect .* [screenRect(3), screenRect(4),
screenRect(3), screenRect(4)];

    % Draw buttons
    Screen('FillRect', window, [55, 116, 207], buttonRep1Rectnorm);
    Screen('FillRect', window, [62, 193, 94], buttonRep2Rectnorm);

    Screen('TextSize', window, 75);
    DrawFormattedText(window, messageRep, 'center', 'center');

    Screen('TextSize', window, 60);
    DrawFormattedText(window, buttonRep1Text, 'center', 'center', [0 0 0], [], [],
[], [], [], buttonRep1Rectnorm);
    DrawFormattedText(window, buttonRep2Text, 'center', 'center', [0 0 0], [], [],
[], [], [], buttonRep2Rectnorm);

    Screen('Flip', window);

    while true

        % If < is pressed on clicker, go back
        [~, ~, keyCode] = KbCheck;
        if keyCode(KbName('PageUp'))
            % Wait until the key is released
            while KbCheck
                % Do nothing, just wait
            end
            % Set movie time to the beginning and replay the video
            Screen('SetMovieTimeIndex', currentKaraoke, 0);
            PsychPortAudio('FillBuffer', audioHandleKar, z');
            PsychPortAudio('Start', audioHandleKar, 1, 0, 1);
            Screen('PlayMovie', currentKaraoke, 1);
            % Display frames until the end of the video
            while true
                [~, ~, keyCode] = KbCheck;
                % Check if a specific key is pressed (e.g., 'ESCAPE' key)
                if keyCode(KbName('space'))
                    % Exit the loop if 'space' key is pressed
                    break
                end
                tex = Screen('GetMovieImage', window, currentKaraoke);
                if tex > 0
                    % Draw the current frame
```



```
                    Screen('DrawTexture', window, tex);
                    Screen('Flip', window);
                    % Close the texture to release resources
                    Screen('Close', tex);
                else
                    % End of the video
                    break
                end
            end
            Screen('PlayMovie', currentKaraoke, 0);
            PsychPortAudio('Stop', audioHandleKar);

            Screen('FillRect', window, [55, 116, 207], buttonRep1Rectnorm);
            Screen('FillRect', window, [62, 193, 94], buttonRep2Rectnorm);

            Screen('TextSize', window, 75);
            DrawFormattedText(window, messageRep, 'center', 'center');

            Screen('TextSize', window, 60);
            DrawFormattedText(window, buttonRep1Text, 'center', 'center', [0 0 0],
[], [], [], [], [], buttonRep1Rectnorm);
            DrawFormattedText(window, buttonRep2Text, 'center', 'center', [0 0 0],
[], [], [], [], [], buttonRep2Rectnorm);
            Screen('Flip', window);

        end
        WaitSecs(0.1);

        % Move on to next stim if > is clicked
        [~, ~, keyCode] = KbCheck;
        if keyCode(KbName('PageDown'))
            % Wait until the key is released
            while KbCheck
                % Do nothing, just wait
            end
            break;
        end
    end
    % Brief pause to release key buffer
    WaitSecs(0.1);

    % Close resources
    delete(audioKar);

    % Video display
    message = ['SING ' ...
        '\n'...
        '\n Press > to continue'];
    Screen('TextSize', window, 75);
    DrawFormattedText(window, message, 'center', 'center');
```



```
    Screen('Flip', window);

    while true
        [~, ~, keyCode] = KbCheck;
        if keyCode(KbName('PageDown'))
            % Wait until the key is released
            while KbCheck
                % Do nothing, just wait
            end
            WaitSecs(0.1);
            break
        end
    end

    % Separate audio and video using ffmepg
    audioIns = 'tempAudioIns.wav';
    ffmpegCommand = sprintf('ffmpeg -i "%s" -vn -acodec pcm_s16le -ar 44100 -ac 2
"%s"', instrumentalVidPath, audioIns);
    [~, ~] = system(['start /B cmd /C "', ffmpegCommand, '" >NUL 2>&1']);

    %Load Instrumental video
    currentInstrumental = Screen('OpenMovie', window, instrumentalVidPath);

    % Set up audio using Psychtoolbox
    [y, ~] = audioread(audioIns);

    % Set up audio recording from the microphone
    record(recorder);

    % Start video playback
    Screen('PlayMovie', currentInstrumental, 1);

    % Start audio playback & recording
    PsychPortAudio('FillBuffer', audioHandleIns, y');
    PsychPortAudio('Start', audioHandleIns, 1, 0, 1);

    while true
        [~, ~, keyCode] = KbCheck;
        % Check if a specific key is pressed (e.g., 'ESCAPE' key)
        if keyCode(KbName('space'))
            % Exit the loop if 'space' key is pressed
            break
        end
        tex = Screen('GetMovieImage', window, currentInstrumental);
        if tex > 0
            % Draw the current frame
            Screen('DrawTexture', window, tex);
            Screen('Flip', window);
            % Close the texture to release resources
            Screen('Close', tex);
```



```
        else
            % End of the video
            break;
        end
    end

    % Stop video and audio (both playback & recording)
    Screen('PlayMovie', currentInstrumental, 0);
    PsychPortAudio('Stop', audioHandleIns);
    stop(recorder);

    % Get the recorded audio data
    audioData = getaudiodata(recorder);
    % Save the audio data to a file
    audiowrite('recordedAudio.wav', audioData, 44100);
    currentAudioPB = 'recordedAudio.wav';

    % Move to recording playback screen
    messagePB = ['WAIT'...
        '\n' ...
        '\n Listen to your recording' ...
        '\n'...
        '\n Press > to continue'];
    Screen('TextSize', window, 75);
    DrawFormattedText(window, messagePB, 'center', 'center');
    Screen('Flip', window);

    % Initialize recording playback
    [f, fs] = audioread(currentAudioPB);

    % Set playback volume
    volumeLevel = 10; % 900% volume increase I think? (1.5% would be 50% volume
increase)
    PsychPortAudio('Volume', audioHandlePB, volumeLevel);

    % Start recording playback & wait for key press to move on
    PsychPortAudio('FillBuffer', audioHandlePB, f');
    PsychPortAudio('Start', audioHandlePB, 1, 0, 1);

    while true
        [~, ~, keyCode] = KbCheck;
        if keyCode(KbName('PageDown'))
            % Wait until the key is released
            while KbCheck
                % Do nothing, just wait
            end
            WaitSecs(0.1);
            break
        end
    end
```

```
    PsychPortAudio('Stop', audioHandlePB);
    delete(currentAudioPB);

    % Record again? screen
    messageR = 'Would you like to... ';
    button1Text = 'LISTEN again, press <';
    button2Text = 'SING again, press X';
    button3Text = 'Continue, press >';

    button1Color = [55, 116, 207];
    button2Color = [247, 231, 86];
    button3Color = [62, 193, 94];

    buttonWidth = 600;
    buttonHeight = 150;

    % Get the screen dimensions
    screenRect = Screen('Rect', window);

    button1Rect = [0.06, 0.7, 0.06 + buttonWidth / screenRect(3), 0.7 +
buttonHeight / screenRect(4)];
    button2Rect = [0.365, 0.7, 0.365 + buttonWidth / screenRect(3), 0.7 +
buttonHeight / screenRect(4)];
    button3Rect = [0.675, 0.7, 0.675 + buttonWidth / screenRect(3), 0.7 +
buttonHeight / screenRect(4)];

    button1Rectnorm = button1Rect .* [screenRect(3), screenRect(4), screenRect(3),
screenRect(4)];
    button2Rectnorm = button2Rect .* [screenRect(3), screenRect(4), screenRect(3),
screenRect(4)];
    button3Rectnorm = button3Rect .* [screenRect(3), screenRect(4), screenRect(3),
screenRect(4)];

    Screen('FillRect', window, button1Color, button1Rectnorm);
    Screen('FillRect', window, button2Color, button2Rectnorm);
    Screen('FillRect', window, button3Color, button3Rectnorm);

    Screen('TextSize', window, 75);
    DrawFormattedText(window, messageR, 'center', 'center');

    Screen('TextSize', window, 55);
    DrawFormattedText(window, button1Text, 'center', 'center', [0 0 0], [], [], [],
[], [], button1Rectnorm);
    DrawFormattedText(window, button2Text, 'center', 'center', [0 0 0], [], [], [],
[], [], button2Rectnorm);
    DrawFormattedText(window, button3Text, 'center', 'center', [0 0 0], [], [], [],
[], [], button3Rectnorm);

    Screen('Flip', window);
```



```
    % Wait for a mouse click on the button

    fileClicks = fopen(stimOrderPath, 'a');

    while true
        % Move on to next stim if > is clicked
        [~, ~, keyCode] = KbCheck;
        if keyCode(KbName('PageDown'))
            % Wait until the key is released
            while KbCheck
                % Do nothing, just wait
            end
            break;
        end
        % Brief pause to release key buffer
        WaitSecs(0.1);


        % Go back to recording screen if Button 2 is clicked on
        % Move back to recording screen if little screen is clicked
        [~, ~, keyCode] = KbCheck;
        if keyCode(KbName('/?'))
            % Wait until the key is released
            while KbCheck
                % Do nothing, just wait
            end

            fprintf(fileClicks, 'Participant clicked re-record on trial %d \n',
trialNum);

            % Set movie time to the beginning and replay the video
            Screen('SetMovieTimeIndex', currentInstrumental, 0);
            Screen('PlayMovie', currentInstrumental, 1);
            record(recorder);
            PsychPortAudio('FillBuffer', audioHandleIns, y');
            PsychPortAudio('Start', audioHandleIns, 1, 0, 1);
            % Display frames until the end of the video
            while true
                [~, ~, keyCode] = KbCheck;
                % Check if a specific key is pressed (e.g., 'ESCAPE' key)
                if keyCode(KbName('space'))
                    % Exit the loop if 'space' key is pressed
                    break
                end
                tex = Screen('GetMovieImage', window, currentInstrumental);
                if tex > 0
                    % Draw the current frame
                    Screen('DrawTexture', window, tex);
                    Screen('Flip', window);
```



```
                    % Close the texture to release resources
                    Screen('Close', tex);
                else
                    break
                end
            end
            Screen('PlayMovie', currentInstrumental, 0);
            PsychPortAudio('Stop', audioHandleIns);
            stop(recorder);

            % Get the recorded audio data
            audioData = getaudiodata(recorder);
            % Save the audio data to a file
            audiowrite('recordedAudio.wav', audioData, 44100);
            currentAudioPB = 'recordedAudio.wav';

            % Move to recording playback screen
            messagePB = ['WAIT'...
                '\n' ...
                '\n Listen to your recording' ...
                '\n'...
                '\n Press > to continue'];
            Screen('TextSize', window, 75);
            DrawFormattedText(window, messagePB, 'center', 'center');
            Screen('Flip', window);

            % Initialize recording playback
            [f, fs] = audioread(currentAudioPB);
            audioHandlePB = PsychPortAudio('Open', [], [], 0, [], 1);

            % Set playback volume
            volumeLevel = 10; % 900% volume increase I think? (1.5% would be 50%
volume increase)
            PsychPortAudio('Volume', audioHandlePB, volumeLevel);

            % Start recording playback
            PsychPortAudio('FillBuffer', audioHandlePB, f');
            PsychPortAudio('Start', audioHandlePB, 1, 0, 1);

            while true
                [~, ~, keyCode] = KbCheck;
                if keyCode(KbName('PageDown'))
                    % Wait until the key is released
                    while KbCheck
                        % Do nothing, just wait
                    end
                    WaitSecs(0.1);
                    break
                end
            end
```



```
            Screen('PlayMovie', currentInstrumental, 0);
            PsychPortAudio('Stop', audioHandleIns);
            PsychPortAudio('Stop', audioHandlePB);

            % Go back to selection screen
            Screen('FillRect', window, button1Color, button1Rectnorm);
            Screen('FillRect', window, button2Color, button2Rectnorm);
            Screen('FillRect', window, button3Color, button3Rectnorm);

            Screen('TextSize', window, 75);
            DrawFormattedText(window, messageR, 'center', 'center');

            Screen('TextSize', window, 60);
            DrawFormattedText(window, button1Text, 'center', 'center', [0 0 0], [],
[], [], [], button1Rectnorm);
            DrawFormattedText(window, button2Text, 'center', 'center', [0 0 0], [],
[], [], [], button2Rectnorm);
            DrawFormattedText(window, button3Text, 'center', 'center', [0 0 0], [],
[], [], [], button3Rectnorm);
            Screen('Flip', window);
        end

        % If < is pressed on clicker, go back to listening screen
        [~, ~, keyCode] = KbCheck;
        if keyCode(KbName('PageUp'))
            % Wait until the key is released
            while KbCheck
                % Do nothing, just wait
            end
            % Set Karaoke movie time to the beginning and replay the video
            Screen('SetMovieTimeIndex', currentKaraoke, 0);
            PsychPortAudio('FillBuffer', audioHandleKar, z');
            PsychPortAudio('Start', audioHandleKar, 1, 0, 1);
            Screen('PlayMovie', currentKaraoke, 1);
            % Display frames until the end of the video
            while true
                [~, ~, keyCode] = KbCheck;
                % Check if a specific key is pressed (e.g., 'ESCAPE' key)
                if keyCode(KbName('space'))
                    % Exit the loop if 'space' key is pressed
                    break
                end
                tex = Screen('GetMovieImage', window, currentKaraoke);
                if tex > 0
                    % Draw the current frame
                    Screen('DrawTexture', window, tex);
                    Screen('Flip', window);
                    % Close the texture to release resources
                    Screen('Close', tex);
```

```
                else
                    % End of the video
                    break
                end
            end
            Screen('PlayMovie', currentKaraoke, 0);
            PsychPortAudio('Stop', audioHandleKar);

            % Go back to selection screen

            Screen('FillRect', window, button1Color, button1Rectnorm);
            Screen('FillRect', window, button2Color, button2Rectnorm);
            Screen('FillRect', window, button3Color, button3Rectnorm);

            Screen('TextSize', window, 75);
            DrawFormattedText(window, messageR, 'center', 'center');

            Screen('TextSize', window, 60);
            DrawFormattedText(window, button1Text, 'center', 'center', [0 0 0], [],
[], [], [], [], button1Rectnorm);
            DrawFormattedText(window, button2Text, 'center', 'center', [0 0 0], [],
[], [], [], [], button2Rectnorm);
            DrawFormattedText(window, button3Text, 'center', 'center', [0 0 0], [],
[], [], [], [], button3Rectnorm);
            Screen('Flip', window);

        end

        WaitSecs(0.01); % Short delay to avoid high CPU usage
    end

    % Close resources
    Screen('CloseMovie', currentInstrumental);
    Screen('CloseMovie', currentKaraoke)
    delete(audioIns);
    delete(currentAudioPB);
    fclose(fileClicks);
```

Break option screen

Add an optional break halfway through experimental script (after trial 21).

```
    % Add a pause after 21 trials ( = halfway through)
    if i ==21

        messagePause = ['Time for a break!' ...
            '\n'...
            '\n Press > to continue with the experiment'];

        Screen('TextSize', window, 75);
        DrawFormattedText(window, messagePause, 'center', 'center');
```



```
        Screen('Flip', window);

        while true
            [~, ~, keyCode] = KbCheck;
            if keyCode(KbName('PageDown'))
                % Wait until the key is released
                while KbCheck
                    % Do nothing, just wait
                end
                WaitSecs(0.1);
                break
            end
        end

    end
```

End of Part 1

```
        % End loop
end

PsychPortAudio('Close', audioHandleIns);
PsychPortAudio('Close', audioHandlePB);
```

Part 2 - Speaking

Starts part 2 of the experiment which displays the same sentences to be read out.

```
message1 = ['This is the end of Part 1' ...
    '\n' ...
    '\n Almost done! Part 2 will be much quicker.' ...
    '\n' ...
    '\n' ...
    '\n' ...
    '\n' ...
    '\n Press > to continue to Part 2.'];

Screen('TextSize', window, 70);
DrawFormattedText(window, message1, 'center', 'center');
Screen('Flip', window);
```

PICTURE

```
while true
    [~, ~, keyCode] = KbCheck;
    if keyCode(KbName('PageDown'))
        % Wait until the key is released
        while KbCheck
```



```
            % Do nothing, just wait
        end
        WaitSecs(0.1);
        break
    end
end

message2 = ['\n PART 2 - INSTRUCTIONS' ...
    '\n'...
    '\n Read the sentences in a normal voice' ...
    '\n as they appear on the screen.' ...
    '\n ' ...
    '\n There will be 21 sentences to read.' ...
    '\n' ...
    '\n Press > to continue. '];

Screen('TextSize', window, 70);
DrawFormattedText(window, message2, 'center', 'center');
Screen('Flip', window);
```

PICTURE

```
while true
    [~, ~, keyCode] = KbCheck;
    if keyCode(KbName('PageDown'))
        % Wait until the key is released
        while KbCheck
            % Do nothing, just wait
        end
        WaitSecs(0.1);
        break
    end
end

% Read sentences from a .txt file & convert into a cell array
sentenceID = fopen('C:\Users\Coline
Caillol\Desktop\PhD\XP_EPG\Stimuli\PsychToolBox\lyrics.txt', 'r');
sentences = textscan(sentenceID, '%s', 'Delimiter', '\n');
fclose(sentenceID);
sentences = sentences{1};

% Shuffle the order of sentences
shuffledIndices = randperm(length(sentences));
shuffledSentences = sentences(shuffledIndices);

rect = Screen('Rect', window);
xPos1 = rect(3) - 1100;
yPos1 = rect(4) - 100;
```

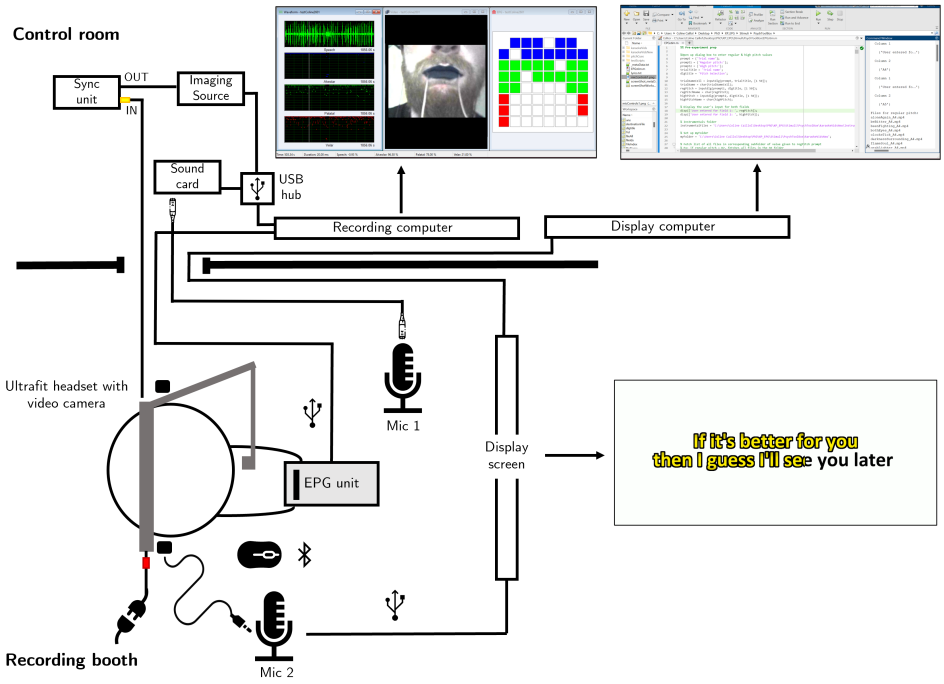


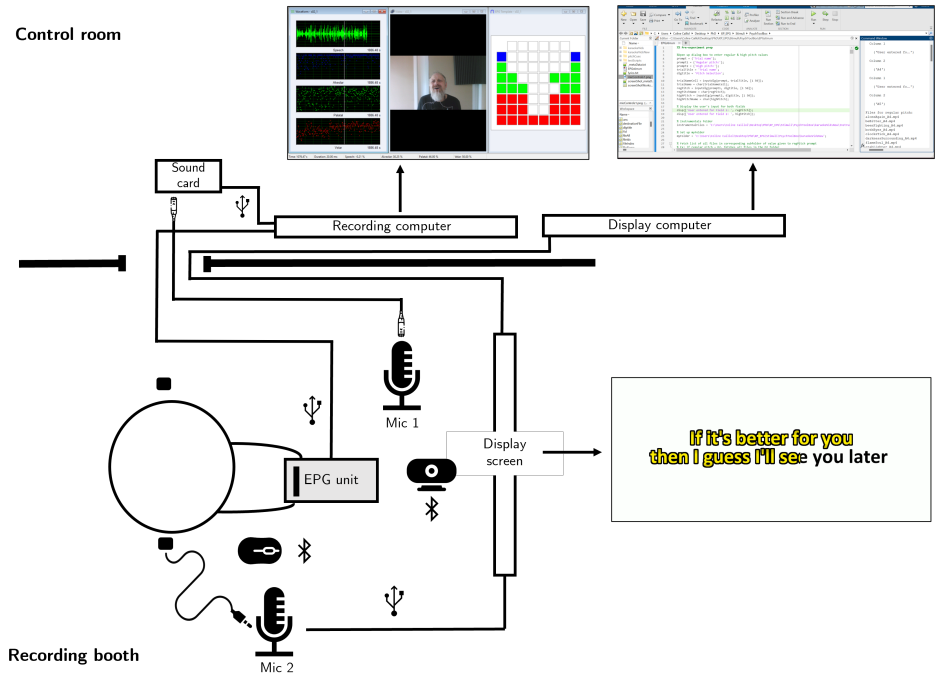
```
% Loop through all shuffled sentences

for i = 1:length(shuffledSentences)
    currentSentence = shuffledSentences{i};
    trialInfo1 = sprintf('Sentence %d of %d', i, length(shuffledSentences));

    Screen('TextSize', window, 70);
    DrawFormattedText(window, currentSentence, 'center', 'center');

    Screen('TextSize', window, 50);
    DrawFormattedText(window, trialInfo1, xPos, yPos);

    messageClick = 'Press > to continue';
    Screen('TextSize', window, 50);
    DrawFormattedText(window, messageClick, xPos1, yPos1)

    Screen('Flip', window);

    while true
        [~, ~, keyCode] = KbCheck;
        if keyCode(KbName('PageDown'))
            % Wait until the key is released
            while KbCheck
                % Do nothing, just wait
            end
            WaitSecs(0.1);
            break
        end
    end

end

% Add sentence order to previously created meta data file

% Open the file in append mode
fileSentence = fopen(stimOrderPath, 'a');
fprintf(fileSentence, '%s\n', 'Reading order:');

for k = 1:length(shuffledSentences)
    sentenceIndex = shuffledSentences(k);
    sentence = char(sentenceIndex);
    fprintf(fileSentence, '%s\n', sentence);
end

fclose(fileSentence);
```

End message

```
messageEnd = ['THE END' ...
    '\n' ...
    '\n Thank you for your participation!'...
    '\n'...
    '\n Press > to exit'];
Screen('TextSize', window, 70);
DrawFormattedText(window, messageEnd, 'center', 'center');

Screen('Flip', window);

while true
    [~, ~, keyCode] = KbCheck;
    if keyCode(KbName('PageDown'))
        % Wait until the key is released
        while KbCheck
            % Do nothing, just wait
        end
        WaitSecs(0.1);
        break
    end
end

sca;
```

B.9 Experiment setups for lip tracking via color segmentation (top) and facial landmark (bottom)





B.10 Experiment consent form

Université Paris Cité

CLILLAC arp

CONSENT FORM

Articulatory Constraints in Singing: An electropalatographic & lip tracking study

This study has been reviewed and received ethics clearance through the Université Paris Cité Research Ethics Committee (n° 00012023-41).

ID Code [to be completed by experimenter]:

Name:

Date of birth:

E-mail address:

Phone number:

I hereby confirm that I have been informed of the purpose of this study, and of the policy of data collection. I had the opportunity to ask questions and my questions have been answered.

I have been informed of the safety measures and health protocols put in place in the laboratory. I agree to respect these rules and regulations.

I understand that my participation is voluntary. I am aware that I have the right to withdraw at any time, without having to provide a reason and without any negative consequences.

I accept to follow the instructions given by the experimenter, before, during and after the experiment. If I do not follow the instructions provided, the experimenter is free to stop my participation at any time.

I agree to take part in this experiment: ☐ Yes ☐ No

I also accept [please check what applies]:

☐ That the data collected from this experiment will be used by the research team for research purposes and submitted to anonymous computerized treatment at the CLILLAC-ARP lab. Research data will be kept for 2 years after the end of the study.

☐ That the results of this study will be shared in academic settings (conferences, articles etc.) both in-person and online, with the understanding that I will not be identified by name.

☐ That the audio recordings of my voice collected during the study be shared in academic settings, with the understanding that I will not be identified by name but that my voice cannot be anonymized and may lead to my identification nevertheless.

☐ That the video recordings of my lips collected during the study be shared in academic settings, with the understanding that although the video will have been modified to only show a pixelized outline of the lips, any other physical feature will have been cropped out, and that I will not be identified by name, it may lead to my identification nevertheless.

By providing my consent,

- I am not releasing the investigator(s) or involved institution(s) from their legal and professional responsibilities.
- I am not waiving my legal rights.
- I am not authorizing any commercial use of the data collected.
- I can change my choices by contacting the experimenter within four weeks after taking part in the experiment.

Signed in .....(location), on .....(date)

Participant signature (preceded by the mention “read and approved”):

Experimenter signature:

2 original copies

CONTACT INFORMATION

Main investigator:

Coline CAILLOL, PhD candidate at Université Paris Cité (CLILLAC-ARP)

coline.caillol@u-paris.fr

06.38.56.89.52

C Production Experiment 2: Electropalatography (EPG)

C.1 Questions for semi-guided interview

Université Paris Cité

CLILLAC a/p

POST-EXPERIMENT PARTICIPANT QUESTIONNAIRE

ID Code [to be completed by the experimenter]:

1) Could you tell what exactly the experiment was about/what we were trying to look at specifically?
2) Do you actively think about your accent/pronunciation when you sing or is it not something you take into consideration?
3) Do you feel like you change your accent/pronunciation when you sing?
    a. If so, why?
    b. Do you think it's linked to the genre of music you sing?
4) Do you actively try to maintain your spoken accent/pronunciation when you sing?
    a. If so, why?
    b. Do you think it's linked to the genre of music you sing?
5) Do you expect singers to sing with a certain accent/pronunciation?
6) Regarding the pronunciation of /t/ specifically (words like later, better, lighter…), have you ever noticed a difference in the way you speak and sing them?
    a. Do you think it is linked to trying to sound American?
    b. Do you think it is linked to whether you're emphasizing a certain word?
    c. Do you think it is linked to ease of singing (higher pitch or faster tempo)?
7) Any other comments or remarks regarding the study?

Observations:

C.2 Abbreviated transcripts of participant interviews

C.2.1 Participant s01

Experimenter: *So do you, could you tell or do you have any inkling or idea what the experiment was about like? What I was actually trying to get at by making you sing.*

s01: Not at all.

But instinctively you knew because you made a comment about it. At some point you asked if you know you should be singing this like you were in a rock band with more of an American accent.

s01: Yeah, yeah, yeah

So I'm basically working on Americanization ...

s01: And that's the, I think that's the challenge when you start thinking about dialects because in Scotland we don't say Ts, yeah. It's a... glottal...?

But at the same time, you wouldn't really sing with a glottal stop, would you?

s01: No, no, there's a band called Del Amitri, I don't know if you know them. They're, they're from Glasgow, yes. And they were very, very popular in the mid 80s - early 90s and they had quite a few songs in the charts and there's quite a few of them where he actually sings *bottles* [with glottal stop] of whatever.

But do you think this is because they sort of insisted on their Scottish identity?

s01: Well no that's the weird thing the weird thing because he sings like you would think he was American. You know if you didn't know that they were Scottish you wouldn't... You wouldn't know that they were Scottish just by listening. It's not like The Proclaimers. Yeah, very Scottish. Yes, in that singing their accent and stuff. There are quite a few Scottish... But Wet Wet Wet was also really, really popular in the early 80s and again from the West Coast of Scotland, and they've got very, very broad dialect.

...

So I was trying to elicit, to get you to do some flaps – which you did – but you do, you do seem to do them in your speech as well. Have you ever noticed that?

s01: Well I've never pronounced Ts [stops] – that's my dialect. But usually it's a lot more or a lot less pronounced you know if I go back home and I speak with my friends but yeah. So *bottle* [glottal] is would be the normal way or *better* [glottal] but even then, because I've traveled so much and you know living in different countries, obviously that influences... When people don't understand what you're saying, it's like OK, the fifth time I have to repeat myself so then you change a bit. But also, because I've done lot of recording as well and one of the things that might have come through that I don't know but Bs, Ps and Ts are not good for recording.

They're not?

s01: No, because it creates a big spike, you get a big boom. So we try, especially with backing vocals you try and get people to actually not pronounce Ps... You know, and it just puts a massive spike on the thing and trying to get rid of it afterwards is difficult. So you just say, can you not say it as powerful, can you use more of a B sound rather than P, then that's wow, much better. And with the S as well. Again, if you've got like 10 people singing.

...

So do you feel like you change the way you pronounce words when you sing?

s01: Yeah, definitely. But nine times out of ten, I'm mimicking stuff that's already... So I would normally try and sing like the person I'm trying to imitate. But on the songs that we've sung ourselves, I don't know, I'd have to go back.

So you don't actively think about it.

s01: No but because I've recorded a lot of stuff that's not, that's not mine. But you know, a friend of mine who's written the... and he's from London and and he sings with an American accent.

When you're listening to music, do you expect to hear a specific type of accent or would you be for example, would you be surprised listening to a Scottish band if they sounded really Scottish, if you didn't know they were.

s01: In the Rock, in the rock genre yeah... I mean there are... There's another Scottish band called Runrig and they're a folk-rock band yeah, so they were folk band, right. So they were singing traditional Scottish songs but with a rock kind of background and one of the vocalists, he sang in his original dialect. But they were from the far north and and it's almost as nice as proper English, you know what I mean? It doesn't really have this dialect that I have. There's also this really popular now, a bit of an overweight young guy, he's got Tourette's? Blonde haired guy from Glasgow and he's got this really sweet voice. Yes, I think he's singing with an American accent, I don't think it's Scottish because when he talks, yeah, in the crowd, you know, when he's live, it's like you can, you can tell he's Scottish and then it's just like sort of dies down when you're singing. And what about Amy Winehouse? She had a really particular, but it's kind of like her voice as well. And then I guess because she was singing, kind of she's American.

I guess it's also the genre, I think, that influences, you know, how you sound. And because she was singing sort of like bluesy...

s01: She does sing with a London accent as well. She does. And some of the some of the stuff that she's got, you know, some of the ways that she pronounces stuff. But there was a period where there was a lot of English girls singers that were singing that way, like Adele. Adele had a lot of this like, almost this London thing coming through. There's a few, a few just in bits and pieces...

C.2.2 Participant s02

Experimenter: *My first question is, is can you sort of tell what the actual experiment was about?*

s02: I think maybe how often the tongue hits the palate. But so there was one, there was one thing I think I was bad at one point. Because at one point I realized I was singing with an American accent and I'm British and I know I should be singing with a British accent.

There you go. You just touched on it. Basically what I was, what I'm looking at is this idea that some British singers, obviously big generalization, but have this sort of tendency to sound a little bit more American when they sing.

s02: Definitely.

And what I'm looking at specifically is the pronunciation of T. So that's why there's so many words later, better, etcetera.

s02: I Americanize myself.

Yes, you Americanize yourself pretty much all the time, I think. So that was really interesting because you're actually the first one to do that. The others have done it, but have mostly kept a British accent when singing.

s02: There's a technical reason for that because I've been doing this for so many years just because I know. That the when you equalize, when you're equalizing the voice, the T the T's are really aggressive. So I've learned to tone my T's down and and my S's too, all of the S's in this specific study. But the T the T's and the S's I because I have so much experience, I've learned to sing more Americanized to turn them down because it's more pleasant in the ear.

[...] I'm looking at articulatory settings and saying that maybe pronouncing a T the American way, so saying better [flap] *instead of better* [stop] *is easier to sing.*

s02: *Better* [stop], *better* [flap]. No, it's not exactly. It's actually the same same difficulty. But it's just nicer on the ear.

OK, so you think it has more to do with the sort of aesthetic?

s02: It's a taste and aesthetic, artistic aesthetic.

OK, because there's this idea that, you know, also if you're singing at a higher pitch, then your mouth is open wider.

s02: That's interesting. I actually had trouble with the with the with the palate thing because there's something on the front of the teeth and I couldn't go because normally when I when I sing I try not to be lazy and and really spread my yeah I couldn't do it exactly how I wanted to.

So you think it's a question of aesthetic, but so you have a sort of pop band and you also have a hardcore band. Do you pronounce your T's or just your overall accent in both those bands is the same?

s02: No, it's not exactly the same because the pop band has lots of rap sections. And I do the rap bits. And so and the rap bits is about getting information across. So I'm I'm actually rapping in a very English with a very English accent. I need to have people understand what

I'm saying. And then the chorus is who cares? It's just like, la, la, rock and, you know, it's just the sing-along bit. So it's much less important. And also it's other people singing along with me.

But so you're saying the British accent in a way. Is more easily understandable than when you're rapping.

s02: Yeah, for me.

So you just adapt it to your own perspective, I guess.

s02: And in the hardcore band, it's quite monotonic. So it's just shouting and screaming.

And do you do you have an idea of which sort of accent, I guess you use more?

s02: Yes, I do, actually. Yeah, I mean, I'm in full English mode. Yeah, I'm in full English mode because it's all about aggression. So I'm getting, I'm going for the aggressive Ts. It's full aggressivity, full Ts. Make sense? And then with the pop one, I am Americanizing myself more.

So you think it is linked, in a way, to the genre of music that you're saying? Your pronunciation and everything?

s02: Yeah, for sure. Hardcore there's a lot of American hardcore bands. So, but the tradition of my specific hardcore band is more British hardcore. So it's all about aggressively saying what you want to say.

And so do you think that also kind of has to do with your... Not inspirations, but sort of the bands that you you take after when you make your music?

s02: Yeah, we all have references. We all have artistic references. It's difficult to say for us specifically because we've been doing it for such a long time and we really, yeah, we listen to such a lot of music all the time. Yeah, I listen from, I listen to Taylor Swift to right up to Stockhauser, John Cage, you know. so I go from the most commercial to the most avant-garde.

...

But yeah the pronunciation thing, it was, it definitely was unconscious. Because I only thought about it halfway through. So totally unconscious. That's also the aesthetic of the music and the poetry, the specific words that make me automatically do that without thinking.

C.2.3 Participant s03

Experimenter: *Well, my first question is basically, did you, could you kind of guess or could you tell what the actual experiment is about?*

s03: I assume you're you're doing some sort of AI singing machine or whatever. That was my first thought. And that you were trying to trying to, you know, get the get the pronunciation right. But then I thought, if that's the case, you just do that from audio. You wouldn't need the, you wouldn't need the contact points.

OK, so I mean, no, I don't really use AI. I do have a thing where I will look at your lip opening. But here the idea is this is really phonetic. So I am looking at articulation...

s03: I think I tended to pronounce all my Ts.

You did. So it was quite interesting because you pronounced, you basically pronounced them all. If I'm correct, I've like picked up on two that you pronounced more American like.

s03: I think a couple of I actually sang more. Because I thought, wait a minute, I'm not singing, I'm just, I'm saying words in tune. And so then I started singing a bit more. But I would have to say, I mean. Wouldn't it depend? I mean, for me, if anyone asked me that question of why, it would depend entirely on who our influences are.

Yes, yes, that's a big, big part of it.

s03: I'm a huge Beatles fanatic, so and I've read. Like my bookshelf, you know, from The Beatles is about that long. And so I've read all the interviews and all the stories and everything. And of course, when they first went to America, what they were asked was why? Why do you sing with an American accent? And you know, well, it's because that's what they've been listening to.But see, the interesting thing is that in the started in the 70s, some British artists started to sing with their own voices. So Ray Davis, The Kinks sounds a lot more English than American. And then you hit the 80s. Have you heard of Billy Bragg?

No.

s03: OK, well, he's an English protest singer to begin with, but he had big chart hits in England. He can't sing brilliantly. He can sing in tune, but he doesn't really have a voice. And one of the main things is he sounds 100% English. Yeah, yeah, yeah. He made no attempt. He's known around the world, but not in America. I don't know.

Well, yeah, the thing is, it's obviously there are some British artists who don't use that sort of American thing, but it's kind of like a standard. So if you do sound English, then it, you know, people are like, oh, that's jarring a little bit.

s03: Like more recent groups like Arctic Monkeys. But there were artists at the end of the 80s, beginning of the 90s, Blur, Damon Alba, London Pulp, Sheffield as well, actually. A lot of the northern bands, Franz Ferdinand. So little bit of a Scottish thing. But actually a lot of, I don't know if you've noticed a lot of the Scottish bands nearly all sound entirely American. So there's also a lot more influenced by the American so that could be a yes.

I think you're right that it's that it's about influences. So do you feel personally like, is this something that you've actively thought about?

s03: Oh, absolutely.

Your pronunciation?

s03: Yeah, because I've tried to get more British.

Right. OK. Interesting.

s03: It annoys me that I speak like this and then I've been listening back to an album that I did because I had to prepare a live show. And so I've been listening back to an album I did just two albums ago. And I was sitting back thinking, wow, I didn't remember that I sang it, that I I Americanized a lot of the a lot of the sounds, a lot of the vowels and the consonants.

And so I made an effort on the on the last album to actually bring it back. Not not exactly how I speak, but just sort of move the dial. Yeah, have a I just wanted it to sound more natural, like like me.

But so you're saying to sound more natural when you're singing or more like you're speaking voice, you are making a conscious effort.

s03: I'm making a conscious effort, yes, to not sing. But as far as I'm concerned, I'm only singing like that because that's how I've heard. That's how I've heard other people's. I wouldn't say that I do it to make because it's easier to sing. It's just that that's the way.

Do you think because you're a little bit like pop folk, I'd say. So do you think also that genre of music specifically sort of, you know, would you have your pronunciation or your accent in mind if you were doing a totally different genre of music?

s03: Like I'm looking at the influence of folk music on you know, folk is very much more sort of you sing, you sound like where you come from. That's the nature of folk, whether it's music or art or whatever. Pop, I think it's up to sort of. I don't think you'll find a general rule. Yeah, that's really broad. And like you said, if you get pop that has a that has a sort of indie edge to it, then, depending on the singer, like you know like Blur or Pulp, or the Arctic Monkeys, they could decide to sort of put their put their regional stamp on it, whereas other bands wouldn't bother.

...

So I listened to a few of your songs just in preparation to see you have one, Everything is better than before –

s03: Yeah, that's the one that's on the film soundtrack at the moment, right. It's from about 3 albums ago.

Because that you sound really American.

s03: Ah yes but that was an "exercice de style" – that's what we call Americana. So like folk with a tiny bit of country. Yeah, I see brought pop slightly into the folk-rock world. I was in Australia and I just wanted to do something different. And I had a bunch of songs and one of them sounded definitely Americana. And it wasn't something I'd done before. I hadn't gone in that direction. And I contacted my producer here and just said, look, I've got all those songs and what don't we do? Can we try and do something that's sort of Americana with a tiny bit of soul or something? And so the album was actually recorded to sound like an album from West Coast America in the 70s. The funny thing is that the album I did after that, which was supposed to be a return to British pop rock. And that's the one I say when I listen back to that I sounded so American, like why am I still sounding dirty? It's obviously a hangover from the previous.

C.2.4 Participant s04

Experimenter: *So that's actually my first question to you is, do you have any sort of idea what I was trying to look at by making you do all these crazy things?*

s04: I don't know. For a while I thought it was for a game or something, like gaming.
Oh, really? Why, What made you think that?
s04: Because it's quite sophisticated and there's lots. It must cost a fortune to set this thing up. And you must have got a grant from somewhere. And I thought, where can she possibly get a grant. It'd probably be somebody like cybernetics. I don't know. But apparently it's not that.
No, it's really something to do with the way people pronounce particular sounds.
s04: Maybe T's and S's or something then because I've noticed there's some and the S's there's lots of them and it's sort of hard to get them all sort of jammed in. But I mean it's doable. Is it something like that?

...

And so that's kind of like my hypothesis and my argument is that in some cases it might just be easier to sing.
s04: It's just easier and you get to get it out. Because with the T, you've got to actually stop the whole machine.
Yes! That's the whole, yes, that's the whole idea.
s04: But if you can possibly stop the machine gracefully and get the T in it, I think personally that it's better. That's better. I sort of make it a point in my music to not sort of mind the American way of of pronouncing things cause right. I think there's a third way as we might say. Yeah, but other people don't care about sort of making a point of I don't like maybe say somebody like Coldplay, they don't care. You know, it's just that's what they want to sing like, you know.
But so you do you think you pay attention to your pronunciation, the articulation of of specific sounds when you sing?
s04: I do. I don't want to sing... I don't want to be... I want to be, to have an individual voice which is not American, right.
So you do take into account the way you pronounce certain things and try to avoid others. That's interesting. And what kind of music, you're a little bit of like post-punk?
s04: Yea, so it's not 1 000 miles from heavy metal. You know, it's rock. Yeah,, it's in the same garden. But I think more heavy metal, they go much more for the American Pronunciation. And I think, I mean between you and me, I probably want to sound like I sort of intellectually got a lot of baggage and I'm clever and all this sort of shit, and so maybe that's the way of doing it. Pronouncing it clearly and articulating and things like that. And let's face it, I'm not sort of Maria Callas or something like that. That's not what it's about sort of thing. I'm more it's just more of a character than sort of virtuosity, right.
Right, I see. Do you think the fact that you're also, so you do say it's about your character... Do you think it's also linked maybe to the genre of music that you sing?
s04: Possibly, yeah. Yeah You're quite right. If you think of the Sex Pistols. Yeah, I think Johnny Rotten. When I heard the first Sex Pistols song in my life, I remember it was a

cultural shock that was gigantic for me. And I thought, oh, you can sing like that to this sort of music. I'm going to want to do that. Then it's a long *chemin* to where we are now. But that, I remember that specifically hearing *Anarchy in the UK* and his particular voice, it was like massive... I mean I'm 65 so previous to that I was like an adolescent. So I wasn't sort of completely like, waking up on that day, either. I'd already listened to tons of music and that's the thing that aggressive rock, that was what was on previously. And they often sing in a sort of Americanized accent. Not heavy. But still there. It's still there. Like Pink Floyd. Yes, it's very slightly Americanized and that sort of thing. And then he's [Johnny Rotten] came on the thing and then loads of other bands as well... I think it's lazy, the American way of doing things. And I mean it's not that I'm anti-American or anything, but I just think it's easy to try and write a song where you're sort of using iconography from America with a lovely Cadillac car and driving in the desert, and New York and... For me, that's an easy way of doing stuff. And I want people to think when they hear me of something much more urban and European.

Yeah, yeah, fair enough.

s04: And more specifically English. And when people hear our music, they say it's just so English what you're doing in the way that it's constructed and the way it's sang as well. And for me, I think, oh, OK, it's job done then.

But so it's interesting that you would say the American way of doing things is easy or well, but do you think you make a conscious effort to sound English in your music, and the default easier version would be to sing with an American accent?

s04: I believe that, yeah.

So it's it's more of a conscious choice and harder in a way to sound British, even if you are British.

s04: Yeah, it is harder, yeah, I think. Because it's not natural, I think.

But why? Because you're English. So why would it not be natural?

s04: I know, because we've just been *abreuvé* with the American stuff, and the origin of the music we're playing is jazz and blues. And that's it's black American. And they've got a specific accent that sort of is just layered on so thick you can't escape as soon as you put something on like. Sort of a Chuck Berry song or something like that. You're there in the Delta of Chicago or somewhere. You know, you know where you are. And I don't want that to happen when people listen to our music. Other people do do that. And so they're evoking this Chicago thing or whatever it is. It's sort of a Bluesy Delta thing, and it's sort of an easy way to evoke something. I want people to evoke something else. Now when you put, for example, a Sonata by Chopin or something like that, of that class, it's going to play the Chopin. When you put something like contemporary 20th century music on, the people who are listening to it have to really sort of strain their ears to culturally get the reference. There is no reference sometimes. No, it's completely not Chopin. It's not Beethoven, Haydn, Chuck Berry or anything. It's slightly different. And I think that that essence of when you hear

music is what I'm really always looking for. So people hear it and go, what's that? And another little thing that I wanted to say now is that a sort of a technique that I use to to make art and also to make music more specifically is to try and have people see one thing that maybe they know, but see it through a glass that is. Completely unfamiliar. Like, for example, to see a Sex Pistols song but seen through a Chopin glass. OK, that would be an effect that I would sort of look for.

C.3 EPG data extraction and reduction MATLAB script

EPG data extraction and reduction

This script extracts EPG activation data for all frames in a specific .csv file and converts them using a lookup table which maps hexadecimal numbers to the EPG palate electrodes. Several masks are computed isolating specific regions of the palate. Activation in percentage is saved to a .txt file, and heatmap and anatomical visualization options can be selected.

Other files needed:

- .csv file with the EPG frames in hexadecimal numbers
- Final .txt file with time stamps and token at least

Initialization

Look-up Table

```
sortVec = [39 23 15 7 60 36 52 12 55 63 47 31 4 20 28 44 30 22 14 6 45 61 37 53 ...
    54 38 62 46 5 13 21 29 26 18 10 2 40 56 32 48 50 34 58 42 0 8 16 24 ...
    27 19 11 3 41 57 33 49 51 35 59 43 1 9 17 25] + 1;
```

EPG regional filters

```
% Filter for alveolar bracing "horseshoe" configuration
filtreAlveoBracing = [
    0 1 1 1 1 1 1 0;
    1 1 1 1 1 1 1 1;
    1 1 0 0 0 0 1 1;
    1 1 0 0 0 0 1 1;
    1 1 0 0 0 0 1 1;
    1 1 0 0 0 0 1 1;
    1 1 0 0 0 0 1 1;
    1 1 0 0 0 0 1 1
];

% Filter for the anterior alveolar region
filtreAnterior = [
    0 1 1 1 1 1 1 0;
    1 1 1 1 1 1 1 1;
    0 0 0 0 0 0 0 0;
    0 0 0 0 0 0 0 0;
    0 0 0 0 0 0 0 0;
    0 0 0 0 0 0 0 0;
    0 0 0 0 0 0 0 0;
    0 0 0 0 0 0 0 0;
];

% Filter for the lateral region
filtreLateral = [
    0 1 0 0 0 0 1 0;
```



```
    1 1 0 0 0 0 1 1;
    1 1 0 0 0 0 1 1;
    1 1 0 0 0 0 1 1;
    1 1 0 0 0 0 1 1;
    1 1 0 0 0 0 1 1;
    1 1 0 0 0 0 1 1;
    1 1 0 0 0 0 1 1;
];
```

Importing raw data

```
% Indicate current participant:
currentParticipant = 's01';

% Open final .txt file
[filename, pathname] = uigetfile({'*.txt*'},'Open final data file');
if ~ischar(filename)
    return;
end
myGrid = readtable(fullfile(pathname, filename),
'FileType','text','VariableNamingRule','preserve');

% Open .csv file with EPG frames in hexadecimal numbers
[filename, pathname] = uigetfile({'*.csv'},'Open the .csv file with EPG frames');
if ~ischar(filename)
    return;
end

myFile = fullfile(pathname, filename);
ops = detectImportOptions(myFile);
ops.VariableTypes = {'double', 'char'};
epgData = readtable(myFile, ops);
epgData.Properties.VariableNames = {'Var1', 'Var2'};

% Convert hexadecimal frames to a binary vector
binaryData = hexToBinaryVector(epgData.Var2);

% Get corresponding frames for boundaries segmented in Praat
% Define frame rate
frameRate = 200;
%Get initial EPG frame
tmin2Frame = round(myGrid.tmin*frameRate);
%Get final frame
tmax2Frame = round(myGrid.tmax*frameRate);
```

Store and clean up data

Initialize storage

```
%Initialize results table
```



```
resultats = cell(size(myGrid,1),1);
```

Apply clean up functions

removeFalseContactsFunction and fixContactlessElectrodesFunction can be found at the end of this script.

```
% Define the threshold for number of isolated grouped electrodes to
% deactivate
n = 3;

% Define the threshold (in %) for adjacent contacts to activate broken electrodes
minPercent = 0.75;

%Loop through time intervals from final .txt file
for idx = 1:size(resultats, 1)
%Initialize data table
    currentDataStore = zeros([8, 8, tmax2Frame(idx)-tmin2Frame(idx)+1]);
    frameCount = 1;
%Loop through EPG frames of current interval
    for jdx = tmin2Frame(idx):tmax2Frame(idx)
        currentFrame = fliplr(binaryData(jdx,:));
        sortedEPG = currentFrame(sortVec);
        frameData = permute(reshape(sortedEPG, [8, 8]), [2, 1]);
        % Filter out false contacts
        frameData = removeFalseContactsFunction(frameData, n);
        % Activate broken electrodes
        frameData = fixContactlessElectrodesFunction(frameData, minPercent);
        currentDataStore(:,:,frameCount) = frameData;
        frameCount = frameCount + 1;
    end
    resultats(idx) = {currentDataStore};
end
```

Data extraction loop

Gets the activation % of corresponding EPG frames for each interval (token) for: whole palate, alveolar + lateral regions and anterior region.

Can also compute visualization in the form of heatmaps and/or anatomical view.

```
resultsTable = table;
myToken = myGrid.text;


for idx = 1:length(myToken)
    currentToken = myToken(idx);
    currentData = resultats(idx);
    concatData = cat(3, currentData{:});

    % Get activation % for each contact
    sumFrame = sum(concatData,3);
```



```
    numFrames = size(concatData, 3);
    activationPercent = 100 * sumFrame / numFrames;

    % To make a heatmap for each token
    imshow(imresize(rescale(activationPercent, 0, 1), 30))
    colormap jet

    % To make an anatomical image for each token
    plotAnatomicalHeatmapFunction(activationPercent, tokenLabel, participantID);

    % Export figures
    exportgraphics(gca, [char(currentParticipant) char(currentToken) '_'
char(currentPitch) '_heatmap.jpg']);
    close

    % Get the overall total activation % of the whole palate for the token
    overallPercent = mean(activationPercent(:));

    % Apply AlveoBracing filter
    filteredActivationValues = activationPercent(filtreAlveoBracing == 1);
    % Compute the mean only over these filtered elements
    overallPercentAlveoBracing= mean(filteredActivationValues);

    % Apply anterior filter
    filteredActivationValues1 = activationPercent(filtreAnterior == 1);
    % Compute the mean only over these filtered elements
    anteriorPercent = mean(filteredActivationValues1);

    resultsTable = [resultsTable; table({currentToken}, overallPercent,
overallPercentAlveoBracing, anteriorPercent,...
                    'VariableNames', {'Token','OverallPercent',
'AlveoBracingPercent', 'AnteriorPercent'})];


end

fileName = [char(currentParticipant) '_activationSummary_speaking.txt'];
writetable(resultsTable, fileName, 'Delimiter', '\t');
```

Additional functions

removeFalseContactsFunction

```
function filteredFrame = removeFalseContactsFunction(frameData, n)
    % Remove false contacts (connected components of 1s of size <= n)
    CC = bwconncomp(frameData);
    numPixels = cellfun(@numel, CC.PixelIdxList);
    for i = find(numPixels <= n)
        frameData(CC.PixelIdxList{i}) = 0;
    end
```

```
    filteredFrame = frameData;
end
```

fixContactlessElectrodesFunction

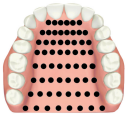
```
function filteredFrame = fixContactlessElectrodesFunction(frameData, minPercent)
    % Activate 0s if >= minPercent of neighbors are 1s (e.g., minPercent = 0.75)

    % Create a copy to modify
    modifiedFrameData = frameData;

    % Loop through each pixel
    for i = 1:numel(frameData)
        if frameData(i) == 0
            [row, col] = ind2sub(size(frameData), i);
            neighborhood = getNeighborhood(frameData, row, col);

            % Count active (1) contacts and total neighbors
            numOnes = sum(neighborhood(:) == 1);
            numTotal = numel(neighborhood);

            % Activate if enough neighbors are active
            if numOnes / numTotal >= minPercent
                modifiedFrameData(i) = 1;
            end
        end
    end

    filteredFrame = modifiedFrameData;
end
```

getNeighborhood

```
function neighborhood = getNeighborhood(frameData, row, col)
    % Get the size of the frameData
    [rows, cols] = size(frameData);

    % Define the range for the neighborhood
    rowRange = max(1, row-1):min(rows, row+1);
    colRange = max(1, col-1):min(cols, col+1);

    % Extract the neighborhood
    neighborhood = frameData(rowRange, colRange);
end
```

plotAnatomicalHeatmapFunction

Requires a .png file of non-activated anatomical palate:



```
function plotAnatomicalHeatmapFunction(activationPercent, tokenLabel, participantID)

    % Load anatomical palate image
    anatImg = imread('anatPalate.png');
    grayImg = rgb2gray(anatImg);
    binaryMask = grayImg < 50;
    cleanMask = bwareaopen(binaryMask, 50);
    props = regionprops(cleanMask, 'Centroid');
    centroids = cat(1, props.Centroid);

    % Sort centroids to match matrix order
    [~, rowIdx] = sort(centroids(:,2));
    centroids = centroids(rowIdx,:);
    rowSizes = [6, 8, 8, 8, 8, 8, 8, 8];
    sortedCentroids = [];
    currentIndex = 1;
    for i = 1:length(rowSizes)
        rowCentroids = centroids(currentIndex:currentIndex + rowSizes(i) - 1, :);
        [~, colIdx] = sort(rowCentroids(:,1));
        sortedCentroids = [sortedCentroids; rowCentroids(colIdx, :)];
        currentIndex = currentIndex + rowSizes(i);
    end

    % Adjust heatmap data to match 8x8 layout
    adjustedData = nan(8, 8);
    adjustedData(1, 2:7) = activationPercent(1, 2:7);
    adjustedData(2:end, :) = activationPercent(2:end, :);

    reshapedData = adjustedData';
    reshapedData = reshapedData(~isnan(reshapedData));

    % Overlay heatmap on anatomical image
    x = sortedCentroids(:, 1);
    y = sortedCentroids(:, 2);
    figure;
    imshow(anatImg); hold on;
    h = scatter(x, y, 100, reshapedData, 'filled');

    % Custom colormap
    colors = hot(4);
    colors(2,:) = [1, 0.5, 0];
    colormap(colors);
```



```
    colormap(flipud(colormap));
```

C.4 Models

Predictor	Estimate	Std. Error	*t* value	*p* value
(Intercept)	24.087	7.379	3.264	$< .05$ *
Pitch (regular)	-1.859	3.613	-0.514	0.6076
Syllable Duration (1/4)	-2.394	3.684	-0.650	0.516
Syllable Duration (1/8)	-1.292	3.456	-0.374	0.709
Pitch (regular):Syllable Duration (1/4)	4.298	5.310	0.809	0.419
Pitch (regular):Syllable Duration (1/8)	-1.579	4.915	-0.321	0.748

*Whole Palate Contact ~ Pitch * Syllable Duration + (1 | Participant)*

Table C.1: *Output of a linear mixed-effects regression model predicting linguopalatal contact for the whole palate as a function of pitch and syllable duration.*

Predictor	Estimate	Std. Error	*t* value	*p* value
(Intercept)	31.455	9.080	3.464	$< .05$ *
Pitch (regular)	-3.503	4.440	-0.789	0.431
Syllable Duration (1/4)	-3.007	4.527	-0.664	0.507
Syllable Duration (1/8)	-1.715	4.246	-0.404	0.686
Pitch (regular):Syllable Duration (1/4)	4.509	6.524	0.691	0.490
Pitch (regular):Syllable Duration (1/8)	-2.459	6.039	-0.407	0.684

*Alveolar Bracing Contact ~ Pitch * Syllable Duration + (1 | Participant)*

Table C.2: *Output of a linear mixed-effects regression model predicting linguopalatal contact for alveolar and lateral regions as a function of pitch and syllable duration.*

Predictor	Estimate	Std. Error	*t* value	*p* value
(Intercept)	18.348	5.890	3.115	$< .05$ *
Pitch (regular)	-0.065	4.718	-0.014	0.989
Syllable Duration (1/4)	-1.784	4.811	-0.371	0.711
Syllable Duration (1/8)	-5.186	4.513	-1.149	0.252
Pitch (regular):Syllable Duration (1/4)	4.758	6.935	0.686	0.493
Pitch (regular):Syllable Duration (1/8)	2.307	6.419	0.359	0.719

*Anterior Contact ~ Pitch * Syllable Duration + (1 | Participant)*

Table C.3: *Output of a linear mixed-effects regression model predicting linguopalatal contact for the anterior region as a function of pitch and syllable duration.*

Predictor	Estimate	Std. Error	*t* value	*p* value
(Intercept)	1.042	0.134	7.746	$< .001$ ***
Pitch (regular)	-0.065	-0.477	-0.514	0.634
Syllable Duration (1/4)	-0.712	0.139	-0.650	0.536
Syllable Duration (1/8)	-0.621	0.130	-0.374	0.536
Pitch (regular):Syllable Duration (1/4)	0.263	0.200	1.311	0.192
Pitch (regular):Syllable Duration (1/8)	-0.086	0.185	-0.463	0.644

*Whole Palate Contact (norm) ~ Pitch * Syllable Duration + (1 | Participant)*

Table C.4: *Output of a linear mixed-effects regression model predicting linguopalatal contact for the whole palate (normalized per participant) as a function of pitch and syllable duration.*

Predictor	Estimate	Std. Error	*t* value	*p* value
(Intercept)	1.096	0.114	9.597	$< .001$ ***
Pitch (regular)	-0.087	-0.624	-0.789	0.534
Syllable Duration (1/4)	-0.448	0.142	-0.664	0.655
Syllable Duration (1/8)	-0.618	0.134	-0.404	0.538
Pitch (regular):Syllable Duration (1/4)	0.233	0.205	1.136	0.258
Pitch (regular):Syllable Duration (1/8)	-0.109	0.190	-0.576	0.565

*Alveolar Bracing Contact (norm) ~ Pitch * Syllable Duration + (1 | Participant)*

Table C.5: *Output of a linear mixed-effects regression model predicting linguopalatal contact for alveolar and lateral regions (normalized per participant) as a function of pitch and syllable duration.*

Predictor	Estimate	Std. Error	*t* value	*p* value
(Intercept)	0.798	0.190	4.195	< .01 **
Pitch (regular)	0.023	0.164	0.142	0.887
Syllable Duration (1/4)	0.018	0.167	0.110	0.912
Syllable Duration (1/8)	-0.178	0.157	-1.133	0.258
Pitch (regular):Syllable Duration (1/4)	0.214	0.241	0.886	0.376
Pitch (regular):Syllable Duration (1/8)	-0.004	0.223	-0.022	0.982

*Anterior Contact (norm) ~ Pitch * Syllable Duration + (1 | Participant)*

Table C.6: *Output of a linear mixed-effects regression model predicting linguopalatal contact for the anterior region (normalized per participant) as a function of pitch and syllable duration.*

Predictor	Estimate	Std. Error	*t* value	*p* value
(Intercept)	23.518	7.629	3.082	< .05 *
Relabeled Pitch (regular)	-0.576	3.877	-0.149	0.882
Syllable Duration (1/4)	-2.683	4.204	-0.638	0.524
Syllable Duration (1/8)	-0.788	3.845	-0.205	0.837
Relabeled Pitch:Syllable Duration (1/4)	4.089	5.539	0.738	0.461
Relabeled Pitch:Syllable Duration (1/8)	-2.381	5.011	-0.475	0.635

*Whole Palate Contact ~ Relabeled Pitch * Syllable Duration + (1 | Participant)*

Table C.7: *Output of a linear mixed-effects regression model predicting linguopalatal contact for the whole palate as a function of k-means determined relabeled pitch and syllable duration.*

Predictor	Estimate	Std. Error	*t* value	*p* value
(Intercept)	30.002	9.418	3.952	< .05 *
Relabeled Pitch (regular)	-0.451	4.775	-0.095	0.924
Syllable Duration (1/4)	-3.066	5.178	-0.592	0.554
Syllable Duration (1/8)	-0.501	4.735	-0.106	0.915
Relabeled Pitch:Syllable Duration (1/4)	-4.455	6.822	0.564	0.573
Relabeled Pitch:Syllable Duration (1/8)	-2.381	6.172	-0.722	0.471

*Alveolar Bracing Contact ~ Relabeled Pitch * Syllable Duration + (1 | Participant)*

Table C.8: *Output of a linear mixed-effects regression model predicting linguopalatal contact for alveolar and lateral regions as a function of k-means determined relabeled pitch and syllable duration.*

Predictor	Estimate	Std. Error	*t* value	*p* value
(Intercept)	18.036	6.209	2.904	< .05 *
Relabeled Pitch (regular)	0.496	5.031	0.099	0.921
Syllable Duration (1/4)	-3.213	5.464	-0.588	0.557
Syllable Duration (1/8)	-4.889	5.004	-0.977	0.330
Relabeled Pitch:Syllable Duration (1/4)	6.643	7.196	0.923	0.357
Relabeled Pitch:Syllable Duration (1/8)	1.542	6.521	0.237	0.813

*Anterior Contact ~ Relabeled Pitch * Syllable Duration + (1 | Participant)*

Table C.9: *Output of a linear mixed-effects regression model predicting linguopalatal contact for the anterior region as a function of k-means determined relabeled pitch and syllable duration.*

Predictor	Estimate	Std. Error	*t* value	*p* value
(Intercept)	0.976	0.143	6.797	< .001 ***
Relabeled Pitch (regular)	0.058	0.145	0.399	0.691
Syllable Duration (1/4)	-0.078	0.158	-0.494	0.622
Syllable Duration (1/8)	-0.072	0.145	-0.495	0.622
Relabeled Pitch:Syllable Duration (1/4)	0.191	0.208	0.918	0.360
Relabeled Pitch:Syllable Duration (1/8)	-0.092	0.189	-0.486	0.627

*Whole Palate Contact (norm) ~ Relabeled Pitch * Syllable Duration + (1 | Participant)*

Table C.10: *Output of a linear mixed-effects regression model predicting linguopalatal contact for the whole palate (normalized per participant) as a function of k-means determined relabeled pitch and syllable duration.*

Predictor	Estimate	Std. Error	*t* value	*p* value
(Intercept)	1.012	0.126	8.024	< .001 ***
Relabeled Pitch (regular)	0.070	0.148	0.479	0.633
Syllable Duration (1/4)	-0.044	0.162	-0.274	0.785
Syllable Duration (1/8)	-0.062	0.149	-0.418	0.677
Relabeled Pitch:Syllable Duration (1/4)	0.169	0.213	0.796	0.427
Relabeled Pitch:Syllable Duration (1/8)	-0.133	0.194	-0.685	0.495

*Alveolar Bracing Contact (norm) ~ Relabeled Pitch * Syllable Duration + (1 | Participant)*

Table C.11: *Output of a linear mixed-effects regression model predicting linguopalatal contact for alveolar and lateral regions (normalized per participant) as a function of k-means determined relabeled pitch and syllable duration.*

Predictor	Estimate	Std. Error	*t* value	*p* value
(Intercept)	0.750	0.198	3.776	$< .01$ **
Relabeled Pitch (regular)	0.100	0.174	0.575	0.566
Syllable Duration (1/4)	-0.013	0.189	-0.070	0.944
Syllable Duration (1/8)	-0.147	0.173	-0.848	0.397
Relabeled Pitch:Syllable Duration (1/4)	0.247	0.249	0.990	0.323
Relabeled Pitch:Syllable Duration (1/8)	-0.052	0.226	-0.232	0.817

Anterior Contact (norm) $\sim$ *Relabeled Pitch * Syllable Duration + (1 | Participant)*

Table C.12: *Output of a linear mixed-effects regression model predicting linguopalatal contact for the anterior region (normalized per participant) as a function of k-means determined relabeled pitch and syllable duration.*

D Production Experiment 3: Magnetic Resonance Imaging (MRI)

D.1 Recruitment flyer

Université Paris Cité

SINGERS WANTED

FOR A LINGUISTICS STUDY ON **THE SINGING VOICE**

If you are:

- British
- Living in Paris
- A singer
- Willing to spend up to an hour in an MRI scanner

We are looking for you to participate in an paid experiment on singing voice pronunciation!



For more information, contact:

Coline Caillol

coline.caillol@etu.u-paris.fr

06 38 56 89 52

You can scan here to email directly:

D.2 MRI DICOM to video MATLAB script

Create videos from rtMRI DICOM files

This script takes the DICOM files resulting from rtMRI acquisitions and converts them into videos.

Import DICOM files and sort

```
% Indicate path where DICOM files are located
folderPath = 'C:\Users\Coline
Caillol\Desktop\IRM\PhD\IRM_2025_07_21_PG\data\SCANS\47000\DICOM\';
dicomFiles = dir(fullfile(folderPath, '*.dcm'));

% Read metadata and sort
numFrames = length(dicomFiles);
infoList = cell(numFrames, 1);
instanceNumbers = zeros(numFrames, 1);

for i = 1:numFrames
    info = dicominfo(fullfile(folderPath, dicomFiles(i).name));
    infoList{i} = info;
    instanceNumbers(i) = info.InstanceNumber;  % Could also try AcquisitionTime
end

[~, sortIdx] = sort(instanceNumbers);
dicomFiles = dicomFiles(sortIdx);
```

Set up the video

```
% Set up video
v = VideoWriter('C:\Users\Coline
Caillol\Desktop\IRM\PhD\IRM_2025_07_21_PG\Vids\sweeterThan.avi', 'Motion JPEG
AVI');  % Try this instead of MPEG-4
v.FrameRate = 10;
open(v);
```

Normalize frame intensities

```
% Compute global intensity range so that there is no flickering
% in the final video
allVals = [];
for i = 1:numFrames
    img = double(dicomread(fullfile(folderPath, dicomFiles(i).name)));
    allVals = [allVals; img(:)];
end
globalMin = min(allVals);
globalMax = max(allVals);
```

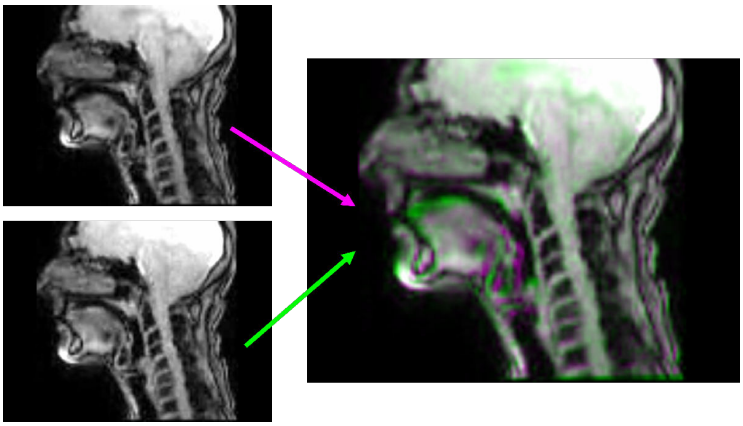
Write video

```
% Write video
for i = 1:numFrames
    img = double(dicomread(fullfile(folderPath, dicomFiles(i).name)));
    img = (img - globalMin) / (globalMax - globalMin);  % Global normalization
    img = im2uint8(img);
    rgbImg = repmat(img, [1 1 3]);  % RGB format
    writeVideo(v, rgbImg);
end

close(v);
```

D.3 Image alignment MATLAB script

Do Image alignment

This script superimposes two similar images and highlights the differences between the two, one in green and the other in magenta. A specific registration function must first be created using the Registration Estimator app to establish a basis of spatial correspondences between two images. It is best to do this using for example the first frame of the videos containing the images meant to be aligned.



Initialization

```
% Path of folder containing the images - modify as needed
cd 'C:\Users\Coline Caillol\Desktop\IRM\PhD\IRM_2025_07_21\finalFrames'

% Read two images
[file1, path1] = uigetfile('*.png', 'Choose spoken T (fixed)');
[file2, path2] = uigetfile('*.png', 'Choose sung T (moving)');

image1 = imread(fullfile(path1, file1));
image2 = imread(fullfile(path2, file2));

% Display the original images
figure;
subplot(2, 2, 1);
imshow(image1);
title('Image 1');

subplot(2, 2, 2);
imshow(image2);
title('Image 2');
```

Quick look to see the superimposed images (before registration):

```
% Superimpose images using imshowpair green = im1 ; magenta = im2
superimposed_image = imshowpair(image1, image2, 'falsecolor');

% Montage with images side by side
superimposed_image = imshowpair(image1, image2, 'montage');
```

Create difference overlay

```
% Make sure the function ie here "registerImages_XX" matches the one created
% through the Registration Estimator app

[MOVINGREG] = registerImages_XX(image2, image1);

figure;
showMatchedFeatures(image1, image2, ...
    MOVINGREG.FixedMatchedFeatures.Location,
MOVINGREG.MovingMatchedFeatures.Location);
legend('Fixed Image', 'Moving Image');

% Create a color overlay of difference
overlayImage = imfuse(image1, MOVINGREG.RegisteredImage);

% Display the overlay image
figure;
imshow(overlayImage);
```

D.4 Distance measures from image MATLAB script

Get distances from image

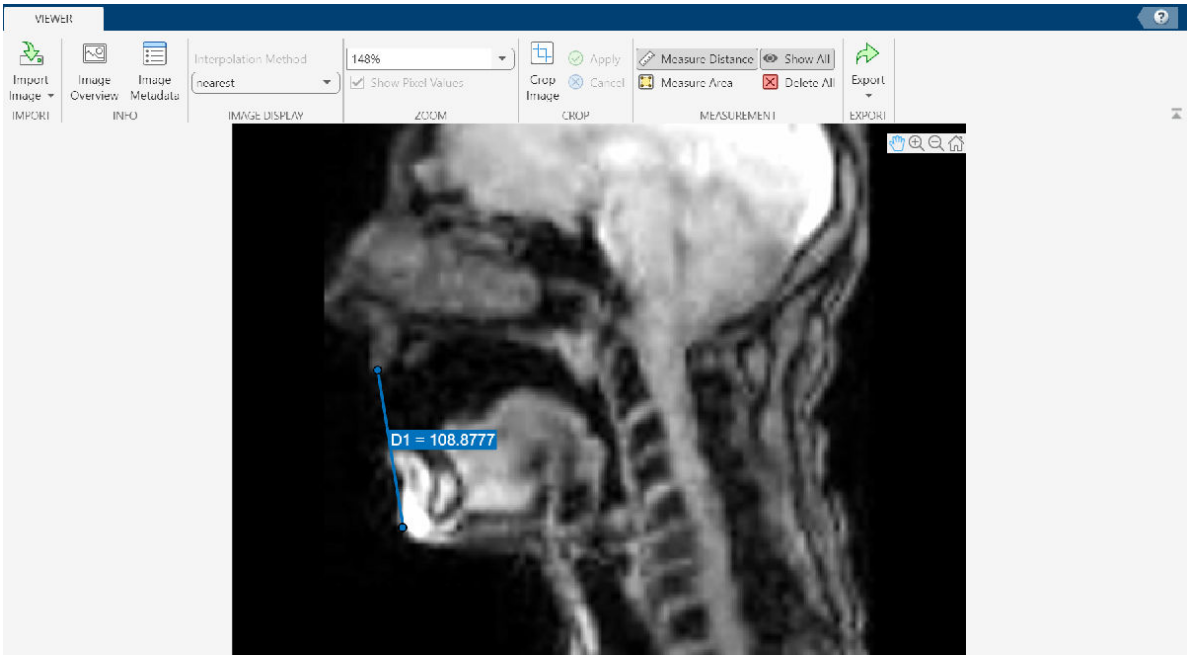
This script lets the user record the euclidean distance in pixels between two points selected manually on an image, for all images in .png or .jpg format located in a single folder.

Initialization

```
% Get all image files in a selected folder

imgDir = uigetdir(pwd, 'Select folder containing frames');

patterns = {'*.png', '*.jpg'};
files = vertcat(...
       dir(fullfile(imgDir, patterns{1})), ...
        dir(fullfile(imgDir, patterns{2})) );

% Set up final table

T = table('Size', [0,6], ...
    'VariableTypes', ["string", 'double', "double","double","double","double"], ...
                  'VariableNames',
["Filename","D1_Value","D1_X1","D1_Y1","D1_X2","D1_Y2"]);
```

Get distance

```
% Set up loop through all images

for k = 1:numel(files)
    imgPath = fullfile(imgDir, files(k).name);
    I = imread(imgPath);

% Open Image Viewer app
imageViewer(I);
```



In the Image Viewer App, use Measure Distance button to draw a line between two points. Then, click export > export measurements and keep the name as D1.

```
% Script waits for D1 variable to appear in the workspace
while ~evalin('base','exist(''D1'',''var'')')
        pause(0.1);
end

% Append measures from D1 to table. This incldues the distance as well as
% coordinates for both points.
S = evalin('base','D1');
    pos = S.Position;
    val = S.Value;


    T = [T; {string(files(k).name), double(val), ...
             double(pos(1)), double(pos(2)), double(pos(3)), double(pos(4))}];

% Clears D1 so that can start again on new image
    evalin('base','clear D1');

 close(gcf);
end

% Write final table as .txt
outTxt = fullfile(imgDir, 'D1_measurements_all.txt');
writetable(T, outTxt, 'Delimiter', '\t');
```

Bibliography

Abercrombie, D. (1957). Direct palatography. *Sprachtypologie und Universalienforschung*, *10*(1), 21.

Acoustica. (2020). *Mixcraft 9 Pro Studio. [Computer software]*. https://acoustica.com/mixcraft

Adorno, T. W. (1990). On popular music. In S. Frith & A. Goodwin (Eds.), *On record: Rock, pop and the written word* (pp. 256–267). Routledge.

Agha, A. (2005). Voice, footing, enregisterment. *Journal of Linguistic Anthropology*, *15*(1), 38–59.

Ahn, S. (2015). Tongue root contributions to voicing in utterance-initial stops in American English. *Proceedings of Meetings on Acoustics*, *25*(1), 1–13.

Alderton, R. (2022). T-tapping in Standard Southern British English: An 'elite' sociolinguistic variant? *Journal of Sociolinguistics*, *26*(2), 287–298.

Anderson, V., Ko, I., O'Grady, W., & Choo, M. (2004). A palatographic investigation of place of articulation in Korean coronal obstruents. *Korean Linguistics*, *12*(1), 1–24.

Androutsopoulos, J. (2014). Mediatization and sociolinguistic change. Key concepts, research traditions, open issues. In J. Androutsopoulos (Ed.), *Mediatization and sociolinguistic change* (pp. 3–48). De Gruyter.

Ansal, L., & Ancy, R. (2021). Research on DNN methods in music source separation tools with emphasis to Spleeter. *International Research Journal on Advanced Science Hub*, *3*(6S).

Anthony, L. (2022). *AntConc. (Version 4.2.0). [Computer software]*. https://www.laurenceanthony.net/software/AntConc

Avid Technology, Inc. (2023). *Sibelius. (Version 2023.5). [Computer software]*. https://www.avid.com/sibelius

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, *47*(1), 31–56.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Beal, J. C. (2009). "You're Not from New York City, You're from Rotherham": Dialect and identity in British indie music. *Journal of English Linguistics*, *37*(3), 223–240.

Bell, A. (1984). Language style as audience design. *Language in Society*, *13*(2), 145–204.

Bell, A. (2002). Back in style: Reworking audience design. In P. Eckert & J. R. Rickford (Eds.), *Style and sociolinguistic variation* (pp. 139–169). Cambridge University Press.

Bell, A., & Gibson, A. (2011). Staging language: An introduction to the sociolinguistics of performance. *Journal of Sociolinguistics*, *15*(5), 555–572.

Belyk, M., Carignan, C., & McGettigan, C. (2024). An open-source toolbox for measuring vocal tract shape from real-time magnetic resonance images. *Behavior Research Methods*, *56*(3), 2623–2635.

Birrer, F. A. J. (1985). Definition and research orientation: Do we need a definition of popular music? In D. Horn (Ed.), *Popular music perspective 2: Papers from the second international conference on popular music studies* (pp. 99–105). International Association for the Study of Popular Music.

Björkner, E. (2008). Musical theater and opera singing—Why so different? A study of subglottal pressure, voice source, and formant frequency characteristics. *Journal of Voice*, *22*(5), 533–540.

Boersma, P., & Weenink, D. (2025). Praat: Doing phonetics by computer (Version 6.4.39). [Computer software]. https://praat.org

Bombien, L., Mooshammer, C., Hoole, P., Kühnert, B., & Schneeberg, J. (2006). An EPG study of initial clusters in varying prosodic conditions. *7th International Seminar on Speech Production (ISSP)*, 35–42.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436. https://doi.org/10.1163/156856897X00357

Braver, A. (2014). Imperceptible incomplete neutralization: Production, non-identifiability, and non-discriminability in American English flapping. *Lingua*, *152*, 24–44.

Bresch, E., & Narayanan, S. (2010). Real-time magnetic resonance imaging investigation of resonance tuning in soprano singing. *The Journal of the Acoustical Society of America*, *128*(5), EL335–EL341.

Burns, P. (1986). Acoustical analysis of the underlying voice differences between two groups of professional singers: Opera and country and western. *The Laryngoscope*, *96*, 549–554.

Burquest, D. A., & Payne, D. L. (1993). *Phonological analysis: A functional approach.* Summer Institute of Linguistics.

Bybee, J. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change*, *14*(3), 261–290.

Caffier, P. P., Ibrahim Nasr, A., Ropero Rendon, M., Wienhausen, S., Forbes, E., Seidner, W., & Nawka, T. (2018). Common vocal effects and partial glottal vibration in professional nonclassical singers. *Journal of Voice*, *32*(3), 340–346.

Caillol, C., & Ferragne, E. (2019). The sociophonetics of British heavy metal music: T-voicing and the FOOT-STRUT split. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences - ICPhS 2019* (pp. 2650–2654).

Caillol, C., & Ferragne, E. (2024). Me[t]al or me[ɾ]al? Factors influencing T flapping in the singing voice [in press]. In A. Przewozny, S. Herment, & J. Rouaud (Eds.), *Realities, communities and norms in spoken English varieties: New perspectives.* Presses Universitaires du Midi.

Campos-Astorkiza, R. (2024). Singing to a genre: Constraints on variable rhoticity in British americana. *Languages*, *9*(6), 203.

Carey, D., & McGettigan, C. (2017). Magnetic resonance imaging of the brain and vocal tract: Applications to the study of speech production and language learning. *Neuropsychologia*, *98*, 201–211.

Carey, J. C., Cohen Jr, M. M., Curry, C. J., Devriendt, K., Holmes, L. B., & Verloes, A. (2009). Elements of morphology: Standard terminology for the lips, mouth, and oral region. *American Journal of Medical Genetics Part A*, *149*(1), 77–92.

Chambers, J. K. (2013). Studying language variation: An informal epistemology. In Chambers, J. K. and Schilling, Nathalie (Ed.), *The handbook of language variation and change* (pp. 1–13). Wiley.

Chambers, J. K., & Trudgill, P. (1998). *Dialectology.* Cambridge University Press.

Chevaillier, G., Guilbault, R., Renard, J.-N., Herman, P., & Huy, P. T. B. (2009). La “voix saturée” du chanteur rock métal, un mécanisme supraglottique performant. *LXVème Congrès, Société Française de Phoniatrie et Pathologies de la Communication*, 331–351.

Chiu, C., & Huang, P.-H. (2023). Lip postures of high vowels in Taiwan Mandarin. In R. Skarnitzl & J. Volín (Eds.), *Proceedings of the 2023 International Congress of Phonetic Sciences - ICPhS 2023* (pp. 1052–1056).

Cho, T., & Keating, P. (2009). Effects of initial position versus prominence in English. *Journal of Phonetics*, *37*(4), 466–485.

Cleland, J., Timmins, C., Wood, S. E., Hardcastle, W. J., & Wishart, J. G. (2009). Electropalatographic therapy for children and young people with Down's syndrome. *Clinical Linguistics & Phonetics*, *23*(12), 926–939.

Collister, L. B., & Huron, D. (2008). Comparison of word intelligibility in spoken and sung phrases. *Empirical Musicology Review*, *3*(3), 109–125.

Cope, A. L. (2016). *Black Sabbath and the rise of heavy metal music*. Routledge.

Coupland, N. (1980). Style-shifting in a Cardiff work-setting. *Language in Society*, *9*(1), 1–12.

Coupland, N. (2007). *Style: Language variation and identity*. Cambridge University Press.

Coupland, N. (2011). Voice, place and genre in popular song performance. *Journal of Sociolinguistics*, *15*(5), 573–602.

Danesi, M. (2007). *Popular culture: Introductory perspectives*. Rowman & Littlefield.

Danesi, M. (2020). Popular culture. *The Year's Work in Critical and Cultural Theory*, *28*(1), 272–289.

Davies, M. (2004). British National Corpus (from Oxford University Press). https://www.english-corpora.org/bnc/

Davies, M. (2008). The Corpus of Contemporary American English (COCA). https://www.english-corpora.org/coca/

De Timmerman, R., De Cuypere, L., & Slembrouck, S. (2024). The globalization of local indexicalities through music: African-American English and the blues. *Journal of Sociolinguistics*, *28*(1), 3–25.

Dehais-Underdown, A., Vignes, P., Crevier-Buchman, L., Demolin, D., Vuissoz, P.-A., Isaieva, K., Fauvel, M., Laprie, Y., & Felblinger, J. (2023). Non-pulmonic initiation in human beatboxing: A real-time MRI study. In R. Skarnitzl & J. Volín (Eds.), *Proceedings of the 20th International Congress of Phonetic Sciences - ICPhS 2023* (pp. 1176–1180).

de Jong, K. J. (1998). Stress-related variation in the articulation of coda alveolar stops: Flapping revisited. *Journal of Phonetics*, *26*(3), 283–310.

de Jong, K. J. (2011). Flapping in American English. In M. van Oostendorp, C. J. Ewen, E. Hume, & K. Rice (Eds.). Wiley.

Derrick, D., & Gick, B. (2011). Individual variation in English flaps and taps: A case of categorical phonetics. *Canadian Journal of Linguistics/Revue canadienne de linguistique*, *56*(3), 307–319.

Dickinson, B. (2017). *What does this button do? Bruce Dickinson: An autobiography*. HarperCollins.

Doblon. (2009). *Karaoke video creator. (Version 2.6.3.6). [Computer software]*. https://www.powerkaraoke.com/

Downing, K. K., & Eglinton, M. (2018). *Heavy duty: Days and nights in Judas Priest*. Hachette Books.

Duncan, D. (2017). Australian singer, American features: Performing authenticity in country music. *Language & Communication*, *52*, 31–44.

Earl, B. (2016). Metal goes 'pop': The explosion of heavy metal into the mainstream. In G. Bayer (Ed.), *Heavy metal music in Britain* (pp. 33–52). Routledge.

Eberhardt, M., & Freeman, K. (2015). 'First things first, I'm the realest': Linguistic appropriation, white privilege, and the hip-hop persona of Iggy Azalea. *Journal of Sociolinguistics*, *19*(3), 303–327.

Echternach, M., Burk, F., Burdumy, M., Traser, L., & Richter, B. (2016). Morphometric differences of vocal tract articulators in different loudness conditions in singing. *PLOS ONE*, *11*(4), e0153792.

Echternach, M., Markl, M., & Richter, B. (2011). Vocal tract configurations in yodelling—prospective comparison of two Swiss yodeller and two non-yodeller subjects. *Logopedics Phoniatrics Vocology*, *36*(3), 109–113.

Echternach, M., Popeil, L., Traser, L., Wienhausen, S., & Richter, B. (2014). Vocal tract shapes in different singing functions used in musical theater singing—a pilot study. *Journal of Voice*, *28*(5), 653–e1.

Echternach, M., Sundberg, J., Arndt, S., Breyer, T., Markl, M., Schumacher, M., & Richter, B. (2008). Vocal tract and register changes analysed by real-time MRI in male professional singers—a pilot study. *Logopedics Phoniatrics Vocology*, *33*(2), 67–73.

Echternach, M., Sundberg, J., Arndt, S., Markl, M., Schumacher, M., & Richter, B. (2010). Vocal tract in female registers—a dynamic real-time MRI study. *Journal of Voice*, *24*(2), 133–139.

Echternach, M., Sundberg, J., Baumann, T., Markl, M., & Richter, B. (2011). Vocal tract area functions and formant frequencies in opera tenors' modal and falsetto registers. *The Journal of the Acoustical Society of America*, *129*(6), 3955–3963.

Eckert, P. (1989). *Jocks and burnouts: Social categories and identity in the high school*. Teachers College Press.

Eckert, P. (2012). Three waves of variation study: The emergence of meaning in the study of sociolinguistic variation. *Annual Review of Anthropology*, *41*(1), 87–100.

Eckert, P. (2018). *Meaning and linguistic variation: The third wave in sociolinguistics*. Cambridge University Press.

Ellis, L. (2009). Articulation characteristics of severely and profoundly deaf children and approaches to therapy: A review of the electropalatography literature. *Language and Linguistics Compass*, *3*(5), 1201–1210.

Epps, J., Smith, J. R., & Wolfe, J. (1997). A novel instrument to measure acoustic resonances of the vocal tract during phonation. *Measurement Science and Technology*, *8*(10), 1112.

Ferragne, E. (2025). *CminR-Praatik. (Version 2.0)*. https://github.com/emmanuelferragne/CminR-Praatik/tree/v2.0

Ferragne, E., & Pellegrino, F. (2010). Formant frequencies of vowels in 13 accents of the British Isles. *Journal of the International Phonetic Association*, *40*(1), 1–34.

Fisher, J., Kayes, G., & Popeil, L. (2019). Pedagogy of different sung genres. In G. F. Welch, D. M. Howard, & J. Nix (Eds.), *The Oxford handbook of singing* (pp. 707–727). Oxford University Press.

Flanagan, P. J. (2019). 'A Certain Romance': Style shifting in the language of Alex Turner in Arctic Monkeys songs 2006–2018. *Language and Literature*, *28*(1), 82–98.

Fletcher, S. G. (1989). Palatometric specification of stop, affricate, and sibilant sounds. *Journal of Speech, Language, and Hearing Research*, *32*(4), 736–748.

Forster, C., & Hardcastle, W. (1998). An electropalatographic (EPG) study of the speech of two stuttering subjects. *International Journal of Language & Communication Disorders*, *33*(sup1), 358–363.

Fougeron, C., & Keating, P. A. (1996). Articulatory strengthening in prosodic domain-initial position. *UCLA Working Papers in Phonetics*, 61–87.

Fougeron, C., Meynadier, Y., & Demolin, D. (2000). 62 vs. 96 electrodes: A comparative analysis of Reading and Kay Elemetrics EPG pseudo-palates. *Seminar on Speech Production: Models and Data-CREST Workshop on Models on Speech Production: Motor Planning and Articulatory Modelling*, 1–4.

Foulkes, P. (2006). Sociophonetics. In K. Brown (Ed.), *Encyclopedia of language and linguistics* (pp. 495–500). Elsevier Press.

Fox, J., & Weisberg, S. (2019). *An R Companion to applied regression*. Sage.

Frith, S. (1988). *Music for pleasure: Essays in the sociology of pop*. Routledge.

Frith, S. (2001). Pop music. In S. Frith & W. Straw (Eds.), *The Cambridge companion to pop and rock* (pp. 93–108). Cambridge University Press.

Frith, S. (2011). Music and identity. In S. Hall & P. du Gay (Eds.), *Questions of cultural identity* (pp. 108–127). Sage.

Fuchs, S., & Perrier, P. (2003). An EMMA/EPG study of voicing contrast correlates in German. In M. J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th international Congress of Phonetic Sciences - ICPhS 2003* (pp. 1057–1060).

Fuchs, S., Perrier, P., Geng, C., & Mooshammer, C. (2006). What role does the palate play in speech motor control?: Insights from tongue kinematics for German alveolar obstruents. In *Speech production* (pp. 149–164). Psychology Press.

Fujimura, O. (1986). Relative invariance of articulatory movements: An iceberg model. In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 226–242). Psychology Press.

Fukaya, T., & Byrd, D. (2005). An articulatory examination of word-final flapping at phrase edges and interiors. *Journal of the International Phonetic Association*, *35*(1), 45–58.

Gibbon, F. (2013). *Bibliography of electropalatographic (EPG) studies in English (1957-2013)*. Articulate Instruments, Ltd. http://www.articulateinstruments.com/EPGrefs.pdf

Gibbon, F., Crampin, L., Hardcastle, W., Nairn, M., Razzell, R., Harvey, L., & Reynolds, B. (1998). CLEFTNET (Scotland): a network for the treatment of cleft palate speech using EPG. *International Journal of Language & Communication Disorders*, *33*(sup1), 44–49.

Gibbon, F., Hardcastle, W., & Nicolaidis, K. (1993). Temporal and spatial aspects of lingual coarticulation in /kl/ sequences: A cross-linguistic investigation. *Language and Speech*, *36*(2-3), 261–277.

Gibbon, F., & Wood, S. (2010). Visual feedback therapy with electropalatography (EPG) for speech sound disorders in children. In L. Williams, S. McLeod, & R. McCauley (Eds.), *Interventions in speech sound disorders* (pp. 509–536). Brookes.

Gibbon, F., Yuen, I., Lee, A., & Adams, L. (2007). Normal adult speakers' tongue palate contact patterns for alveolar oral and nasal stops. *Advances in Speech Language Pathology*, *9*(1), 82–89.

Gibson, A. (2019). *Sociophonetics of popular music: Insights from corpus analysis and speech perception experiments* [Doctoral dissertation]. University of Canterbury.

Gibson, A. (2023). Pop Song English as a supralocal norm. *Language in Society*, *53*(3), 1–28.

Gibson, A., & Bell, A. (2012). Popular music singing as referee design. In J. M. Hernandez-Campoy & J. A. Cutillas-Espinosa (Eds.), *Style-shifting in public: new perspectives on stylistic variation* (pp. 139–164). Benjamins.

Gibson, A., & Hay, J. (2025). Perceptual style-shifting across singing and speech: Music activates Pop Song English for NZ listeners. *Language and Speech*.

Gick, B., Wilson, I., & Derrick, D. (2013). *Articulatory phonetics*. Wiley-Blackwell.

Giles, H. (1973). Accent mobility: A model and some data. *Anthropological Linguistics*, *15*(2), 87–105.

Giles, H. (1980). Accommodation theory: Some new directions. *York Papers in Linguistics*, *9*, 105–136.

Giles, H., Mulac, A., Bradac, J. J., & Johnson, P. (1987). Speech accommodation theory: The first decade and beyond. *Annals of the International Communication Association*, *10*(1), 13–48.

Godfrey, J. J., Holliman, E. C., & McDaniel, J. (1992). SWITCHBOARD: Telephone speech corpus for research and development. *Acoustics, Speech, and Signal Processing, IEE Computer Society International Conference*, *1*, 517–520.

Goodwin, A. (1988). Sample and hold: Pop music in the digital age of reproduction. *Critical Quarterly*, *30*(3).

Gramming, P., Sundberg, J., Ternström, S., Leanderson, R., & Perkins, W. H. (1988). Relationship between changes in voice pitch and loudness. *Journal of Voice*, *2*(2), 118–126.

Gray, H., & Lewis, W. H. (1918). *Anatomy of the human body* (20th ed.) [Online edition published May 2000 by Bartleby.com]. Lea & Febiger. https://www.bartleby.com/107/

Gregory, S. W. J., & Webster, S. (1996). A nonverbal signal in voices of interview partners effectively predicts communication accommodation and social status perceptions. *Journal of Personality and Social Psychology*, *70*(6), 1231–40.

Gumperz, J. J. (1982). *Discourse strategies*. Cambridge University Press.

Guzman, M., Acevedo, K., Leiva, F., Ortiz, V., Hormazabal, N., & Quezada, C. (2019). Aerodynamic characteristics of growl voice and reinforced falsetto in metal singing. *Journal of Voice*, *33*(5), 803.e7–803.e13.

Guzman, M., Barros, M., Espinoza, F., Herrera, A., Parra, D., & Lloyd, A. (2014). Resonance strategies revealed in rock singers during production of high notes. *Journal of Singing*, *71*(2).

Halford, R. (2020). *Confess: The autobiography*. Da Capo Press.

Halliday, M. A. (1978). *Language as social semiotic: The social interpretation of language and meaning*. Arnold.

Halo, M. (2007). Interview with Def Leppard: America's Brits [Accessed: 2025-09-04]. *The Aquarian Weekly*. https://www.theaquarian.com/2007/08/08/def-leppard-americas-brits/

Hardcastle, W., Gibbon, F., & Nicolaidis, K. (1991). EPG data reduction methods and their implication for studies of lingual coarticulation. *Journal of Phonetics*, *19*, 251–266.

Hardcastle, W., & Gibbon, F. (2005). Electropalatography as a research and clinical tool: 30 years on. In W. Hardcastle & J. M. Beck (Eds.), *A figure of speech: A festschrift for John Laver* (pp. 39–60). Lawrence Erlbaum.

Hardcastle, W., Gibbon, F., & Jones, W. (1991). Visual display of tongue-palate contact: Electropalatography in the assessment and remediation of speech disorders. *British Journal of Disorders of Communication*, *26*(1), 41–74.

Hardcastle, W., Jones, W., Knight, C., Trudgeon, A., & Calder, G. (1989). New developments in electropalatography: A state-of-the-art report. *Clinical Linguistics & Phonetics*, *3*(1), 1–38.

Harrison, L. M. (2010). Factory music: How the industrial geography and working-class environment of post-war Birmingham fostered the birth of heavy metal. *Journal of Social History*, *44*(1), 145–158.

Hatch, D., & Millward, S. (1987). *From blues to rock: An analytical history of pop music*. Manchester University Press.

Hawkins, P. (1984). *Introducing phonology*. Hutchinson & Co.

Hayes, E. N. (2023). Meeting in the middle: Sociophonetic convergence of Bad Bunny and J Balvin's coda /s/ in their artistic performance speech. *Languages*, *8*(4), 287.

Hennequin, R., Khlif, A., Voituret, F., & Moussallam, M. (2020). Spleeter: A fast and efficient music source separation tool with pre-trained models. *Journal of Open Source Software*, *5*(50), 2154.

Henrich, N., Smith, J., & Wolfe, J. (2011). Vocal tract resonances in singing: Strategies used by sopranos, altos, tenors, and baritones. *The Journal of the Acoustical Society of America*, *129*(2), 1024–1035.

Herd, W., Jongman, A., & Sereno, J. (2010). An acoustic and perceptual analysis of /t/ and /d/ flaps in American English. *Journal of Phonetics*, *38*(4), 504–516.

Hoit, J. D., Jenks, C. L., Watson, P. J., & Cleveland, T. F. (1996). Respiratory function during speaking and singing in professional country singers. *Journal of Voice*, *10*(1), 39–49.

Hollien, H. (1983). The puzzle of the singer's formant. In D. M. Bless & J. H. Abbs (Eds.), *Vocal fold physiology: Contemporary research and clinical issues* (pp. 368–378). College-Hill Press.

Iverson, G. K., & Ahn, S.-C. (2007). English voicing in dimensional theory. *Language Sciences*, *29*(2-3), 247–269.

Jansen, L. (2018). "Britpop Is a Thing, Damn It": On British attitudes toward American English and an americanized singing style. In V. Werner (Ed.), *The Language of Pop Culture* (pp. 116–135). Routledge.

Jansen, L. (2022). *English rock and pop performances*. Benjamins.

Jansen, L., & Westphal, M. (2017). Rihanna works her multivocal pop persona: A morpho-syntactic and accent analysis of Rihanna's singing style. *English Today*, *33*(2), 46–55.

Jespersen, O. (1922). *Language: Its nature, development and origin*. Allen & Unwin.

Johansson, C., Sundberg, J., & Wilbrand, H. (1982). X-Ray study of articulation and formant frequencies in two female singers. *Speech Transmission Laboratory – Quarterly Progress and Status Report*, *23*(4), 117–134.

Johnson, K. R. (2015). *phonR: Tools for Phoneticians and Phonologists. (Version 1.0-7). [Computer software]*. https://CRAN.R-project.org/package=phonR

Joliveau, E., Smith, J., & Wolfe, J. (2004). Tuning of vocal tract resonance by sopranos. *Nature*, *427*(6970), 116.

Kartynnik, Y., Kekec, S., Grishchenko, I., & Grundmann, M. (2019). *Real-time facial surface geometry from monocular video on mobile GPUs*. arXiv: 1907.06724 [cs.CV]. https://arxiv.org/abs/1907.06724

Kayes, G. (2019). Structure and function of the singing voice. In G. F. Welch, D. M. Howard, & J. Nix (Eds.), *The Oxford handbook of singing* (pp. 3–29). Oxford University Press.

Kelly, S., Main, A., Manley, G., & McLean, C. (2000). Electropalatography and the LinguaGraph system. *Medical Engineering & Physics*, *22*(1), 47–58.

Kenyon, J. (1936). *American pronunciation*. University of Michigan Press.

King, H., & Ferragne, E. (2020). Loose lips and tongue tips: The central role of the /r/-typical labial gesture in Anglo-English. *Journal of Phonetics*, *80*, 100978.

Klatt, D. H. (1975). Voice onset time, frication, and aspiration in word-initial consonant clusters. *Journal of Speech and Hearing Research*, *18*(4), 686–706.

Kob, M., Henrich, N., Herzel, H., Howard, D., Tokuda, I., & Wolfe, J. (2011). Analysing and understanding the singing voice: Recent progress and open questions. *Current Bioinformatics*, *6*(3), 362–374.

Kochetov, A. (2020). Research methods in articulatory phonetics I: Introduction and studying oral gestures. *Language and Linguistics Compass*, *14*(4), 1–29.

Kochetov, A. (2025). Linguopalatal contact differences between japanese geminates and singletons across different places and manners. *Journal of the International Phonetic Association*, 1–30.

Kochetov, A., & Colantoni, L. (2011). Coronal place contrasts in Argentine and Cuban Spanish: An electropalatographic study. *Journal of the International Phonetic Association*, *41*(3), 313–342.

Kochetov, A., Colantoni, L., & Steele, J. (2021). Variable assimilation of English word-final /n/: Electropalatographic evidence. *English Language & Linguistics*, *25*(4), 687–718.

Konert-Panek, M. (2017a). Americanisation versus Cockney: Stylisation in Amy Winehouse's singing accent. In V. Kennedy & M. Gadpaille (Eds.), *Ethnic and cultural identity in music and song lyrics* (pp. 77–94). Cambridge Scholars.

Konert-Panek, M. (2017b). Overshooting Americanisation. Accent stylisation in pop singing – Acoustic properties of the BATH and TRAP vowels in focus. *Research in Language*, *15*(4), 371–384.

Konert-Panek, M. (2018). Singing accent Americanisation in the light of frequency effects: LOT unrounding and PRICE monophthongisation in focus. *Research in Language*, *16*(2), 155–168.

Krause, M., & Smith, J. (2017). 'I Stole It from a Letter, off Your Tongue It Rolled.' The performance of dialect in Glasgow's indie music scene. In C. Montgomery & E. Moore (Eds.), *Language and a sense of place: Studies in language and region* (pp. 215–233). Cambridge University Press.

Krause, P., Pili, R. J., Hunt, E., & Pili, R. J. (2024). A Process for measuring lip kinematics using participants' webcams during linguistic experiments conducted online. *Laboratory Phonology*, *15*(1), 1–38.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). LmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*, 1–26.

Kydd, W. L., & Belt, D. A. (1964). Continuous palatography. *Journal of Speech and Hearing Disorders*, *29*(4), 489–492.

Labov, W. (1963). The social motivation of a sound change. *Word*, *19*(3), 273–309.

Labov, W. (1966). *The social stratification of English in New York City*. Center for Applied Linguistics.

Labov, W. (1972). *Sociolinguistic Patterns*. University of Pennsylvania Press.

Ladefoged, P. (1957). Use of palatography. *Journal of Speech and Hearing Disorders*, *22*(5), 764–774.

Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Blackwell.

Lallouache, M. T. (1991). *Un poste "visage-parole" couleur: Acquisition et traitement automatique des contours des lèvres* [Doctoral dissertation]. Grenoble INPG.

Laver, J. (1994). *Principles of phonetics*. Cambridge University Press.

Le Page, R. B. (1968). Problems of description in multilingual communities. *Transactions of the Philological Society*, *67*(1), 189–212.

Le Page, R. B. (1986). Acts of identity. *English Today*, *2*(4), 21–24.

Le Page, R. B. (2017). The evolution of a sociolinguistic theory of language. In F. Coulmas (Ed.), *The handbook of sociolinguistics* (pp. 13–32). Wiley.

Le Page, R. B., Christie, P., Jurdant, B., Weekes, A. J., & Tabouret-Keller, A. (1974). Further report on the sociolinguistic survey of multilingual communities: Survey of Cayo District, British Honduras. *Language in Society*, *3*(1), 1–32.

Le Page, R. B., & Tabouret-Keller, A. (1985). *Acts of identity: Creole-based approaches to language and ethnicity*. Cambridge University Press.

Lee, A., Gibbon, F., & Oebels, J. (2015). Lateral bracing of the tongue during the onset phase of alveolar stops: An EPG study. *Clinical Linguistics & Phonetics*, *29*(3), 236–245.

Lee, A., Liker, M., Fujiwara, Y., Yamamoto, I., Takei, Y., & Gibbon, F. (2022). EPG research and therapy: Further developments. *Clinical Linguistics & Phonetics*, *37*(8), 701–721.

Liker, M., & Gibbon, F. E. (2008). Tongue palate contact patterns of velar stops in normal adult English speakers. *Clinical Linguistics & Phonetics*, *22*(2), 137–148.

Lim, Y., Toutios, A., Bliesener, Y., Tian, Y., Lingala, S. G., Vaz, C., Sorensen, T., Oh, M., Harper, S., Chen, W., et al. (2021). A multispeaker dataset of raw and reconstructed speech production real-time MRI video and 3D volumetric images. *Scientific Data*, *8*(1), 187.

Lindblom, B. (1983). Economy of speech gestures. In P. F. MacNeilage (Ed.), *The production of speech* (pp. 217–245). Springer.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modelling* (pp. 403–439). Springer.

Lindblom, B., & Sundberg, J. (1971). Acoustical consequences of lip, tongue, jaw, and larynx movement. *The Journal of the Acoustical Society of America*, *50*(4B), 1166–1179.

Lindblom, B., & Sundberg, J. (2007). The human voice in speech and singing. In T. Rossing (Ed.), *Handbook of acoustics* (pp. 669–712). Springer.

Liu, Y., Tong, F., de Boer, G., & Gick, B. (2023). Lateral tongue bracing as a universal postural basis for speech. *Journal of the International Phonetic Association*, *53*(3), 712–727.

LoVetri, J. L., & Means Weekly, E. (2003). Contemporary commercial music (CCM) survey: Who's teaching what in nonclassical music. *Journal of Voice*, *17*(2), 207–215.

Lugaresi, C., Tang, J., Nash, H., McGuire, M., Corke, M., Zambaldi, V., Han, H., Steiner, B., Hammoud, R., Todorovic, S., & Hannun, A. (2019). *MediaPipe: A Framework for Building Perception Pipelines*. Retrieved July 4, 2025, from https://github.com/google/mediapipe

Marshall, E., Stuart-Smith, J., Butt, J., & Dean, T. (2024). Variation and change over time in Standard British Choral Accent (1925-2019). *Laboratory Phonology*, *15*(1), 1–39.

Mat Zin, S., Md Rasib, S., Suhaimi, F. M., & Mariatti, M. (2021). The technology of tongue and hard palate contact detection: A review. *Biomedical Engineering Online*, *20*, 1–19.

Mathis, A., Mamidanna, P., Cury, K. M., Abe, T., Murthy, V. N., Mathis, M. W., & Bethge, M. (2018). Deeplabcut: Markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience*, *21*(9), 1281–1289.

Max Planck Institute for Psycholinguistics. (2022). *ELAN. (Version 6.4). [Computer software]*. https://archive.mpi.nl/tla/elan

McAuliffe, M. J., & Cornwell, P. L. (2008). Intervention for lateral /s/ using electropalatography (EPG) biofeedback and an intensive motor learning approach: A case report. *International Journal of Language & Communication Disorders*, *43*(2), 219–229.

McAuliffe, M. J., Lin, E., Robb, M. P., & Murdoch, B. E. (2008). Influence of a standard electropalatography artificial palate upon articulation: A preliminary study. *Folia Phoniatrica et Logopaedica*, *60*(1), 45–53.

McAuliffe, M. J., Ward, E. C., & Murdoch, B. E. (2007). Intra-participant variability in Parkinson's disease: An electropalatographic examination of articulation. *Advances in Speech Language Pathology*, *9*(1), 13–19.

McLennan, C. T., Luce, P. A., & Charles-Luce, J. (2005). Representation of lexical form: Evidence from studies of sublexical ambiguity. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(6), 1308.

McLeod, S., & Searl, J. (2006). Adaptation to an electropalatograph palate: Acoustic, impressionistic, and perceptual data. *American Journal of Speech-Language Pathology*, *15*(2), 192–206.

Meireles, A. R. (2016). Perceptual and acoustic study of voice quality in high-pitched heavy metal singing. *Proceedings of the 8th International Conference on Speech Prosody*, 1245–1249.

Mellinkoff, D. (2004). *The language of the law*. Resource Publications.

Mesiä, S., & Ribaldini, P. (2015). Heavy metal vocals. A terminology compendium. In T.-M. Karjalainen & K. Kärki (Eds.), *Modern heavy metal: Markets, practices and culture* (pp. 383–392).

Middleton, R. (1990). *Studying popular music*. Open University Press.

Milroy, L. (1987). *Observing and analysing natural language: A critical account of sociolinguistic method*. Blackwell.

Milroy, L. (1980). *Language and social networks*. Blackwell.

Moen, I., & Simonsen, H. G. (1997). Effects of voicing on /t, d/ tongue/palate contact in English and Norwegian. *Eurospeech*, 2399–2402.

Moore, R. M. (2016). The unmaking of the English working class: Deindustrialization, reification and the origins of heavy metal. In G. Bayer (Ed.), *Heavy metal music in Britain* (pp. 143–160). Routledge.

Mooshammer, C., Geumann, A., Hoole, P., Alfonso, P., van Lieshout, P. H., & Fuchs, S. (2003). Coordination of lingual and mandibular gestures for different manners of articulation. In M. J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences - ICPhS 2003* (pp. 81–84).

Mooshammer, C., Hoole, P., & Geumann, A. (2007). Jaw and order. *Language and Speech*, *50*(2), 145–176.

Morrissey, F. A. (2008). Liverpool to Louisiana in one lyrical line: Style choice in British rock, pop and folk singing. In J. A. Fishman, M. A. Locher, & J. Strässler (Eds.), *Standards and norms in the English language* (pp. 195–220). De Gruyter.

Mu-Sonic, Ltd. (2013). *Tophonetics.com*. Retrieved August 29, 2025, from https://tophonetics.com/

Negus, K. (2011). *Producing pop: Culture and conflict in the popular music industry*. Arnold.

Niebuhr, O. (2016). Rich reduction: Sound-segment residuals and the encoding of communicative functions along the hypo-hyper scale. *7th Tutorial & Research Workshop on Experimental Linguistics*, 11–24.

O'Connor, J. D. (1991). *Phonetics*. Penguin Books.

O'Hanlon, R. (2006). Australian hip hop: A sociolinguistic investigation. *Australian Journal of Linguistics*, *26*(2), 193–209.

Osbourne, O. (2010). *I am Ozzy*. Grand Central Publishing.

Owens, K. (2023). Musical genres exhibit distinct sociophonetic targets: An analysis of Quebec French. *Proceedings of the Linguistic Society of America*, *8*(1), 5490.

Patterson, D., & Connine, C. M. (2001). Variant frequency in flap production: A corpus analysis of variant frequency in American English flap production. *Phonetica*, *58*(4), 254–275.

Phillips, W., & Cogan, B. (2009). *Encyclopedia of heavy metal music*. Bloomsbury Publishing.

Pillot-Loiseau, C. (2006). Représentation cognitive de l'efficacité vocale dans la parole et le chant lyrique: Conséquences en rééducation. *Voix Parlée et Chantée (Bilans, Rééducations, Recherche, Travail Vocal, Comédiens et Chanteurs)*, 429–451.

Podesva, R. J. (2007). Phonation type as a stylistic variable: The use of falsetto in constructing a persona. *Journal of Sociolinguistics*, *11*(4), 478–504.

Popeil, L. (2007). The multiplicity of belting. *Journal of Singing*, *64*(1), 77–80.

Potter, J. (1998). *Vocal authority*. Cambridge University Press.

Pouplier, M., Hoole, P., & Scobbie, J. M. (2011). Investigating the asymmetry of English sibilant assimilation: Acoustic and EPG data. *Laboratory Phonology*, *2*(1), 1–33.

Puri, S. (2014). The Trooper or the Sandman? Iron Maiden's conservatism versus Metallica's experimentalism in their philosophies towards musical creativity. *Metal Music Studies*, *1*(1), 69–85.

Puterbaugh, P. (1983). Anglomania: The Second British Invasion [Magazine article]. *Rolling Stone*. https://www.rollingstone.com/music/music-news/new-wave-1980s-second-british-invasion-52016/3/

R Core Team. (2024). *R: A Language and Environment for Statistical Computing. (Version 4.4.0). [Computer software]*. https://www.R-project.org/

Ramsammy, M. (2023). Contextual reduction of word-final /l/ in Spanish: An EPG study. *Journal of the International Phonetic Association*, *53*(2), 333–360.

Recasens, D. (1991). On the production characteristics of apicoalveolar taps and trills. *Journal of Phonetics*, *19*(3-4), 267–280.

Recasens, D. (2007). Flapping in unconstrained alveolars. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences - ICPhS 2007* (pp. 389–392).

Recasens, D., & Pallarès, M. D. (2000). A study of F1 coarticulation in VCV sequences. *Journal of Speech, Language, and Hearing Research*, *43*(2), 501–512.

RStudio Team. (2025). *RStudio: Integrated Development Environment for R. (Version 2025.05.1+513). [Computer software]*. https://posit.co/download/rstudio/

Sakakibara, K.-I., Fuks, L., Imagawa, H., & Tayama, N. (2004). Growl voice in ethnic and pop styles. *Proceedings of the International Symposium on Musical Acoustics, ISMA 2004*, 1–4.

Schilling-Estes, N. (1998). Investigating "self-conscious" speech: The performance register in Ocracoke English. *Language in Society*, *27*(1), 53–83.

Schilling-Estes, N. (2004). Investigating stylistic variation. In J. K. Chambers, P. Trudgill, & N. Schilling-Estes (Eds.), *The handbook of language variation and change* (pp. 375–401). Wiley.

Scotto Di Carlo, N. (2007a). Effect of multifactorial constraints on intelligibility of opera singing (I). *Journal of Singing*, *63*(4), 1–13.

Scotto Di Carlo, N. (2007b). Effect of multifactorial constraints on intelligibility of opera singing (II). *Journal of Singing*, *63*(5), 559–567.

Sergeant, D. (2019). The developing voice. In G. F. Welch, D. M. Howard, & J. Nix (Eds.), *The Oxford handbook of singing* (pp. 189–223). Oxford University Press.

Shibata, S. (1968). A study of dynamic palatography. *Annual Bulletin, Research Institute of Logopedics and Phoniatrics, University of Tokyo*, *2*, 28–36.

Shockey, L. (2003). *Sound patterns of spoken English*. Blackwell.

Silverstein, M. (2003). Indexical order and the dialectics of sociolinguistic life. *Language & Communication*, *23*(3-4), 193–222.

Simpson, P. (1999). Language, culture and identity: With (another) look at accents in pop and rock singing. *Multilingua*, *18*(4), 343–368.

Smialek, E., Depalle, P., & Brackett, D. (2012). A spectrographic analysis of vocal techniques in extreme metal for musicological analysis. *Non-Cochlear Sound: Proceedings of the 38th International Computer Music Conference, ICMC 2012*, 88–93.

Sparrow, W. A., & Newell, K. M. (1998). Metabolic energy expenditure and the regulation of movement economy. *Psychonomic Bulletin & Review*, *5*, 173–196.

Spracklen, K. (2017). 'Yours is the Earth and everything that's in it, and–which is more–you'll be a man, my son': Myths of British masculinity and Britishness in the construction and reception of Iron Maiden. *Metal Music Studies*, *3*(3), 405–419.

Spreafico, L., Pucher, M., & Matosova, A. (2018). UltraFit: A speaker-friendly headset for ultrasound recordings in speech science. *Proceedings of Interspeech 2018*, 1517–1520.

Steele, J., Colantoni, L., & Kochetov, A. (2019). Gradient assimilation in French cross-word /n/ + velar stop sequences. *Journal of the International Phonetic Association*, *49*(1), 87–108.

Stephenson, L. (2003). An EPG study of repetition and lexical frequency effects in alveolar to velar assimilation. In M. J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences - ICPhS 2003* (pp. 1891–1894).

Steriade, D. (2000). Paradigm uniformity and the phonetics-phonology boundary. *Papers in Laboratory Phonology*, *5*, 313–334.

Stone, M., Faber, A., Raphael, L. J., & Shawker, T. H. (1992). Cross-sectional tongue shape and linguopalatal contact patterns in [s], [ʃ], [f], and [l]. *Journal of Phonetics*, *20*(2), 253–270.

Stone, M., & Hamlet, S. (1982). Variations in jaw and tongue gestures observed during the production of unstressed/d/s and flaps. *Journal of Phonetics*, *10*(4), 401–415.

Stone, R. E., Jr., Cleveland, T. F., & Sundberg, P. J. (2003). Acoustic and aerodynamic characteristics of country-western, operatic and broadway singing styles compared to speech. *The Journal of the Acoustical Society of America*, *113*(4_Supplement), 2242–2243.

Storey, J. (2021). *Cultural theory and popular culture: An introduction*. Routledge.

Straw, W. (1993). Popular music and postmodernism in the 1980s. In S. Frith, A. Goodwin, & L. Grossberg (Eds.), *Sound and vision: The music video reader* (pp. 3–21). Routledge.

Stuart-Smith, J., Pryce, G., Timmins, C., & Gunter, B. (2013). Television can also be a factor in language change: Evidence from an urban dialect. *Language*, *89*(3), 501–536.

Sundberg, J. (1970). Formant structure and articulation of spoken and sung vowels. *Folia Phoniatrica et Logopaedica*, *22*(1), 28–48.

Sundberg, J. (1974). Articulatory interpretation of the "singing formant". *The Journal of the Acoustical Society of America*, *55*(4), 838–844.

Sundberg, J. (1987). *The science of the singing voice*. Northern Illinois University Press.

Sundberg, J. (1990). What's so special about singers? *Journal of Voice*, *4*(2), 107–119.

Sundberg, J. (1995). The singer's formant revisited. *Voice*, *4*, 106–119.

Sundberg, J. (2009). Articulatory configuration and pitch in a classically trained soprano singer. *Journal of Voice*, *23*(5), 546–551.

Sundberg, J., Cleveland, T. F., Stone, R., & Iwarsson, J. (1999). Voice source characteristics in six premier country singers. *Journal of Voice*, *13*(2), 168–183.

Sundberg, J., & Rothenberg, M. (1986). Some phonatory characteristics of singers and nonsingers. *Speech Transmission Laboratory – Quarterly Progress and Status Report*, *4*, 65–77.

Sundberg, J., & Thalén, M. (2015). Respiratory and acoustical differences between belt and neutral style of singing. *Journal of Voice*, *29*(4), 418–425.

Sundberg, J., Thalén, M., & Popeil, L. (2012). Substyles of belting: Phonatory and resonatory characteristics. *Journal of Voice*, *26*(1), 44–50.

Sung, K., & Kochetov, A. (2018). Allophonic variation in English coronal stops: An EPG corpus study. *Toronto Working Papers in Linguistics*, *40*.

Tabain, M. (2012). Jaw movement and coronal stop spectra in Central Arrernte. *Journal of Phonetics*, *40*(4), 551–567.

Tabain, M., & Beare, R. (2022). Trans-vocalic coronal consonant coarticulation in Central Arrernte: An electro-palatographic study. *Journal of the International Phonetic Association*, *52*(3), 389–420.

Tagliamonte, S. A. (2006). The linguistic variable. In *Analysing Sociolinguistic Variation* (pp. 70–98). Cambridge University Press.

Tagliamonte, S. A. (2011). *Variationist sociolinguistics: change, observation, interpretation*. Wiley.

Thalen, M., & Sundberg, J. (2001). Describing different styles of singing: A comparison of a female singer's voice source in "classical", "pop", "jazz" and "blues". *Logopedics Phoniatrics Vocology*, *26*(2), 82–93.

The MathWorks Inc. (2024). *MATLAB. (Version 2024a). [Computer software]*. https://www.mathworks.com/products/matlab.html

Titze, I. R. (1997). Noise in the voice. *Journal of Singing*, *53*(5), 40–47.

Tollfree, L. (1999). South East London English: Discrete versus continuous modelling of consonantal reduction. In P. Foulkes & G. J. Docherty (Eds.), *Urban voices: Accent studies across the British Isles* (pp. 163–184). Routledge.

Traser, L., Burdumy, M., Richter, B., Vicari, M., & Echternach, M. (2013). The effect of supine and upright position on vocal tract configurations during singing—a comparative study in professional tenors. *Journal of Voice*, *27*(2), 141–148.

Trudgill, P. (1983). Acts of conflicting identity: The sociolinguistics of British pop-song pronunciation. In *On dialect: Social and geographical perspectives* (pp. 141–160). Blackwell.

Trudgill, P. (1986). *Dialects in contact*. Blackwell.

Trudgill, P. (2000). *Sociolinguistics: An introduction to language and society*. Penguin Books.

Trudgill, P. (2014). Diffusion, drift, and the irrelevance of media influence. *Journal of Sociolinguistics*, *18*(2), 213–222.

Umeda, N. (1977). Consonant duration in American English. *Journal of the Acoustical Society of America*, *61*, 846–858.

Vaux, B. (2000). Flapping in English. *Linguistic Society of America*, *7*, 1–11.

Verhoeven, J., Mariën, P., De Clerck, I., Daems, L., Reyes-Aldasoro, C., & Miller, N. (2019). Asymmetries in speech articulation as reflected on palatograms: A meta-study. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences - ICPhS 2019* (pp. 2821–2825).

Walser, R. (2015). *Running with the devil: Power, gender, and madness in heavy metal music*. Wesleyan University Press.

Warner, N., Fountain, A., & Tucker, B. V. (2009). Cues to perception of reduced flaps. *The Journal of the Acoustical Society of America*, *125*(5), 3317–3327.

Warner, N., & Tucker, B. V. (2011). Phonetic variability of stops and flaps in spontaneous and careful speech. *The Journal of the Acoustical Society of America*, *130*(3), 1606–1617.

Warner, N., & Tucker, B. V. (2017). An effect of flaps on the fourth formant in English. *Journal of the International Phonetic Association*, *47*(1), 1–15.

Weinstein, D. (2000). *Heavy metal: The music and its culture*. Da Capo Press.

Weinstock, J. (2021). *Pop culture for beginners*. Broadview Press.

Wells, J. C. (1982a). *Accents of English 1: An introduction*. Cambridge University Press.

Wells, J. C. (1982b). *Accents of English 2: The British Isles*. Cambridge University Press.

Wells, J. C. (2000). *Longman pronunciation dictionary* (2nd ed.). Pearson ESL.

Werner, V. (2018). Linguistics and pop culture. Setting the scene(s). In V. Werner (Ed.), *The language of pop culture* (pp. 4–26). Routledge.

Werner, V. (2022). *Pop cultural linguistics*. Oxford Research Encyclopedia of Linguistics. Retrieved September 21, 2025, from https://oxfordre.com/linguistics/view/10.1093/acrefore/9780199384655.001.0001/acrefore-9780199384655-e-999

Westbury, J. R., & Keating, P. A. (1986). On the naturalness of stop consonant voicing. *Journal of Linguistics*, *22*(1), 145–166.

Westphal, M. (2018). Pop culture and the globalization of non-standard varieties of English: Jamaican Creole in German reggae subculture. In V. Werner (Ed.), *The language of pop culture* (pp. 95–115). Routledge.

Westphal, M., & Jansen, L. (2021). English in global pop music. In B. Schneider, T. Heyd, & M. Saraceni (Eds.), *Bloomsbury world Englishes volume 1: Paradigms* (pp. 190–206). Bloomsbury Publishing.

Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C. O., Woo, K., Yutani, H., & Dunnington, D. (2025). *ggplot2. (Version 3.5.2). [Computer software]*. https://ggplot2.tidyverse.org

Wrench, A. (2007). Advances in EPG palate design. *Advances in Speech Language Pathology*, *9*(1), 3–12.

Wrench, A., & Balch-Tomes, J. (2022). Beyond the edge: Markerless pose estimation of speech articulators from ultrasound and camera images using DeepLabCut. *Sensors*, *22*(3), 1133.

Zangger-Borch, D., & Sundberg, J. (2011). Some phonatory and resonatory characteristics of the rock, pop, soul, and swedish dance band styles of singing. *Journal of Voice*, *25*(5), 532–537.

Zangger-Borch, D., Sundberg, J., Lindestad, P.-Å., & Thalén, M. (2004). Vocal fold vibration and voice source aperiodicity in 'dist' tones: A study of a timbral ornament in rock singing. *Logopedics Phoniatrics Vocology*, *29*(4), 147–153.

Zhang, Q. (2005). A Chinese yuppie in Beijing: Phonological variation and the construction of a new professional identity. *Language in Society*, *34*(3), 431–466.

Zue, V. W., & Laferriere, M. (1979). Acoustic study of medial /t, d/ in American English. *The Journal of the Acoustical Society of America*, *66*(4), 1039–1050.

Index

LOT unrounding, 24, 51, 109, 111, 119, 122, 142
TRAP-BATH split, 24, 109, 111, 112, 117, 122, 142

Acts of Identity, 15, 17, 25, 31, 92, 95, 241
Attention to Speech Model, 11
Audience Design, 13, 27, 32, 88, 94, 104

contemporary commercial music, 45, 148, 225

economy of speech gestures, 39, 41, 51, 62, 65–68, 121, 137, 178, 218
EGG, 66
electropalatography (EPG), 55, 147, 169, 170, 172, 173, 178, 179, 190, 200, 233
electropalatography EPG, 174
EMA, 61
enregisterment, 10, 22, 27, 68

flap/flapping, 58, 61–64, 66–68, 174, 176, 178, 195–197, 199, 212, 241

H&H Theory, 40, 62, 137
heavy metal, 77, 78, 86, 103, 107, 110, 121
hypercorrection, 16, 31, 105

indexicality, 8, 22, 27, 53, 68

lateral tongue bracing, 54, 66, 174
lip tracking, 147, 149–151, 157, 169, 179, 223, 225

MRI, 168, 169, 223, 225, 227, 231, 233, 237, 241

overshoot, 31, 54, 66, 105, 200

pop culture, 17–19, 22, 23, 124
Pop Song English, 35, 37, 68, 221
popular music, 19, 22, 23, 26, 28, 37, 45, 78, 95, 103, 110, 112, 123, 223

singability, 37, 50, 51, 65
Smooth Signal Redundancy Hypothesis, 40
sociolinguistic variable, 8, 25, 53, 68, 80
sonority, 50, 66
Speech Accommodation Theory, 12, 31
stop

alveolar stop, 54, 62, 64, 174, 195, 212
glottal stop, 55, 62, 64
style, 10–12, 22–24, 26, 31, 77, 86, 91, 103, 112, 124, 138, 225

tap/tapping, 58, 176
Three Waves of variation study, 8, 22

USA-5 model, 24, 26, 31, 51–53, 63, 77, 109, 111, 124