THESE DOCTORAT d'AIX-MARSEILLE UNIVERSITE

Ecole Doctorale:
Sciences de la Vie et de la Santé (ED 62)

Spécialité :
Neurosciences

Sous la direction de Dr. Daniele Schön

Présentée par Nia CASON
Pour obtenir le grade de DOCTEUR d'AIX-MARSEILLE UNIVERSITE

Sujet de la thèse :

The effect of musical rhythm on spoken language
L'effet du rythme musical sur la parole

Soutenance prévue le 9 décembre, 2013
Devant le jury composé de :

Pr. Sonja KOTZ	University of Manchester	(Rapporteur)
Pr. Simone DALLA BELLA	University of Montpellier	(Rapporteur)
Dr. Corine ASTSANO	University of Toulouse II	(Examinatrice)
Dr. Daniele SCHÖN	University Aix-Marseille & INSERM	(Directeur)

THESE DOCTORAT d'AIX-MARSEILLE UNIVERSITE

Ecole Doctorale:
Sciences de la Vie et de la Santé (ED 62)

Spécialité :
Neurosciences

Sous la direction de Dr. Daniele Schön

Présentée par Nia CASON
Pour obtenir le grade de DOCTEUR d'AIX-MARSEILLE UNIVERSITE

Sujet de la thèse :

The effect of musical rhythm on spoken language
L'effet du rythme musical sur la parole

Soutenance prévue le 9 décembre, 2013
Devant le jury composé de :

Pr. Sonja KOTZ	University of Manchester	(Rapporteur)
Pr. Simone DALLA BELLA	University of Montpellier	(Rapporteur)
Dr. Corine ASTSANO	University of Toulouse II	(Examinatrice)
Dr. Daniele SCHÖN	University Aix-Marseille & INSERM	(Directeur)

TITRE en français: **L'effet du rythme musical sur la parole**

RESUME en français :

La musique et la parole reposent sur une organisation temporelle. En effet, l'anticipation, l'organisation et le groupement temporel y sont nécessaires. Il serait donc possible que des processus domaine-généraux sous-tendent ces deux processus temporels soient mis en jeu.

Pour tester cette hypothèse, trois expériences utilisant des mesures comportementales et électrophysiologiques (EEG) ont été menées afin de déterminer si la perception et la production de la parole peuvent bénéficier d'un amorçage rythmique (i.e. la présentation au préalable d'un rythme musical qui peut renseigner sur les structures temporelles de la parole). En utilisant ces mesures, nous avons montré que le traitement phonologique de pseudo-mots parlés est renforcé lorsque la parole est conforme aux prédictions temporelles des auditeurs (Cason & Schön, 2012). Le traitement phonologique des phrases peut également être amélioré et cet effet d'amorçage augmenté grâce à un entrainement avec les rythmes musicaux (Cason, Astésano & Schön, soumis). Dans une troisième étude, nous avons montré que l'amorçage rythmique peut augmenter la production phonologique chez les enfants sourds (Cason, Hidalgo & Schön, soumis).

Ces trois études montrent que la régularité du rythme musical (plus important que dans la parole) semblerait permettre, de manière générale, la formation de prédictions temporelles précises et une trace mnésique également améliorer la production de la parole chez les enfants souffrants de troubles auditif.

TITRE en anglais: **The effect of musical rhythm on spoken language**

RESUME en anglais:

Music and speech are both reliant on how events occur in time. Both require anticipation about when and what events will occur as well as a temporal and hierarchical organisation of salient and less salient events. These may rely on common, domain-general processes.

With this in mind, three experiments using behavioural and electrophysiological (EEG) measures were conducted which aimed to investigate whether speech perception and production can benefit from rhythmic priming (inducing temporal expectations through music, and which can inform a listener about temporal structures in speech). We have found that phonological processing of spoken pseudowords is enhanced when speech conforms to listener expectations, as measured by behavioural (reaction time) and EEG data (Cason & Schön, 2012). Phonological processing of sentences can also be enhanced via rhythmic priming (behavioural measures) and this priming effect is augmented through training with the musical rhythms (Cason, Astésano & Schön, submitted).

Overall, it seems that the regularity of musical rhythm (over speech rhythm) allows a listener to form precise temporal expectations and a metrical memory trace which can impact on phonological processing of words and sentences, and that rhythmic priming can also enhance articulation performance in hearing-impaired children, perhaps via an enhanced phonological perception.

MOTS-CLES/KEYWORDS
Parole, Rythme, Musique, Prédictions temporelles / Speech, Rhythm, Music, Temporal prediction

ACKNOWLEDGEMENTS

This Ph.D was completed at the INSERM U 1106, Institut de neurosciences des systèmes, Aix-Marseille University II, Marseille, France. It is part of the ‘Europe BRain and MUSic (EBRAMUS) Network’ and was supported by Initial Training Network Marie Curie of the 7th Framework programme of the European Union, grant no. 238157. Secondments and collaborations were carried out in assication with the Reid School of Music, University of Edinburgh, the Department of Experimental Psychology, Oxford University, and the Centre d’Action Médico Sociale Précoce (C.A.M.S.P), La Timone Hospital, Marseille. To these organizations, businesses, institutes and centres I give my thanks.

I would like to express my sincere gratitude to the members of the jury who have accepted to evaluate this work, in particular to the two reviewers Prof. Sonja Kotz and Prof. Simone Dalla Bella, as well as to member of the jury Dr. Corine Astésano.

I am indebted to my supervisor Daniele Schön for his support and guidance during the course of this Ph.D. I offer my profound thanks for this opportunity.

I would like to thank Dr. Katie Overy (University of Edinburgh) and her research group for hosting me during a secondment period. For the same reasons, I would also like to thank Dr. Kate Watkins and Emily Connally (Oxford University). I am especially grateful to Emily Connally for her kind welcome, support and interest in this research during my stay in Oxford.

I would like to acknowledge Céline Hidalgo (C.A.M.S.P) who spent a huge amount of time collecting and analysing data. I would also like to thank Dr. Corine Astésano for her support and expertise in the making of two experiments.

The research performed during the course of this thesis would of course not have been possible without the participants who have taken part in the experimental tasks, to whom I am also grateful.

I would also like to express my thanks and wishes to fellow students of the EBRAMUS network: Tatiana Selchenkova, Floris van Vugt, Marieke van der Steen, Eleanor Harding, Charles-Etienne Benoit, Nicolas Farrugia, Laura Verga, Laura Ferreri, Richard Penninger, Sujin Hong and Irene Alonso Fernández. I also wish to thank PIs of this network, in particular Dr. Barbara Tillman and Dr. Katie Overy.

Finally, I give thanks to friends and my family John, Angela, Rob and Lara who have been supportive for the duration of this thesis, especially to Tania Anconina Papoutsi and Andrew Meso, and all those who have learned how to use skype on my behalf, also to Matt Timelord Banks and J. Algernon Gubay for never forgetting to remind me what music actually is.

TABLE OF CONTENTS

Chapter 1

INTRODUCTION: Music and Speech

1.1 What is interesting about studying music and speech?

Both music and speech can be considered as two universal forms of human communication, which may have similar evolutionary origins (review McDermott & Hauser, 2005; Wallin, Merker & Brown, 2001). Aside from sharing general 'communicative' function, it is true that the functions of music and speech are different; whilst speech (spoken language) is able to convey explicit meaning, the subjective nature of music interpretation makes its functions less obvious. It has thus been viewed in terms of its role in social cohesion (Cross, 2001; Huron, 2001), emotional regulation (Juslin & Sloboda, 2001), or even as simply a by-product of properties of the auditory system (Pinker, 1997). Considering that their functions differ so greatly, why then is it interesting to study these two domains in parallel? Though music and speech may serve different communicative purposes, structural and acoustic similarities between the two have inspired researchers to investigate whether the underlying neural bases of music and speech processing may also show similarities. Not only are basic acoustic features such as pitch and timing important in both domains, but they also share what might be described as more cognitive processes; both music and speech are communicated via a hierarchical, rule-based transmission of auditory events. For instance, Generative Grammar, the study of structural rules governing grammatical formulation, has been used to describe the syntactic rules underlying both language (Chomsky, 1956) and music (Baroni, Maguire & Drabkin, 1983; Lerdahl & Jackendoff, 1983).

In their desire to understand everything, humans take things apart. In linguistics, theoretical grammar identifies the essential components of language: phonology, morphology, syntax, semantics and pragmatics. Each of these components is interactive, with both bottom-up inputs (i.e. semantics is dependent on a coherent syntax, which is in turn dependent on lexical retrieval, which is in turn dependent on phoneme perception), as well as top-down inputs whereby higher levels of processing can impact on lower ones (McClelland, 1987; Langacker, 1987). For instance, phonemic restauration occurs under conditions of degraded sensory input, so that even when information is missing we hear it as occurring as continuously (Samuel, 1987) (notably also a phenomenon observed in music (Bregman, Colantonio & Ahad, 1999)). Similarly, music has been decomposed into the elements of pitch, rhythm, melody and harmonic relations which integrate to form the global percept of music.

Breaking down speech and music into their constituent parts like this may be informative when studying within domains. However, whilst some rough analogies can be made between these components of language and music (e.g. both have a grammatical structure), it may be more informative to compare the cognitive operations which are required for both music and speech processing (Besson & Schön, 2001) (e.g. the processing of grammatical structure in both domains may rely on a process of structural integration (Patel, 1998, 2012)). From this angle, the comparative study of music and language may provide insights into the nature of auditory cognition, and cognition in general.

Whilst relatively little is known about the extent to which cognitive mechanisms may be shared in music and speech - probably because it is still unclear as to what cognitive processes occur within domains - there is robust evidence to show that initial, sensory information encoding is shared (e.g. Patel., 2011; Patel & Iversen, 2007; Wong et al., 2007). Evidently, both music and speech perception rely on the early subcortical encoding of basic acoustic parameters such as pitch, duration and amplitude. At a cognitive processing level, both music and speech rely on the integration of incoming information with an existing context, as well as with existing knowledge/representations. One example of this may be the hierarchical organisation and segmentation of sounds into groups, be it into words in speech or rhythmic clusters in music: this requires not only an online integration of incoming information with preceding sound events, but may also be influenced by experience-based metrical representations.

Cognitive functions in speech and music processing are also dependent on the initial detection of acoustic features of sound - indeed, it has been proposed that early auditory perception is an important input which provides a ‘scaffold’ for the development of general cognitive sequencing abilities (Conway, Pisoni & Kronenberger, 2009). In speech, for example, the phonological perception is required for word recognition. In the same way, perception of pitch events is required for melody perception in music. However, cognitive processes may also be independent of, or may even feed in a top-down manner onto sensory processes, as is the case with phonemic restauration (speech is heard as continuous despite degraded sensory information), or subjective rhythmisation (auditory events being heard as more or less stressed despite the fact they provide identical auditory information), for example.

Many studies have investigated the effect of music on cognitive processes in a broad sense. These have looked at the impact of musical training on what can be said to be general cognitive abilities, such as mathematical ability, second language learning or IQ (e.g. Schellenberg, 2004, 2011). Whilst identifying transfer effects of music to such functions may be interesting, it does not define which cognitive processes musical training may enhance in order to exert this effect (for example, musical training could enhance mnemonic, attentional, and/or representational processes, all of which could potentially enhance cognitive functioning in general). These ‘transfer’ effects also do not necessarily imply a causal effect of musical training.

Considering that brain processing is optimised for efficiency, it seems there would be no advantage in having speech- or music-specific processing systems (Kluender & Greeberg, 1989). From this perspective, it may be possible to identify common processes to both domains, and thus the potential for music and speech processing to interact. In his resource-sharing framework, Patel (Patel, 1998, 2012) proposes a theoretical account for the possibility that cognitive resources for structural integration may be shared by speech and music. Though this process of structural integration may serve different functions in the two domains (syntax/harmonic relations), the neural resources underlying this process may be shared. He thus makes a distinction between these resource networks (engaged during online processing) from representational networks (stored domain-specific knowledge). In the same way that structural integration may be required for syntactic/harmonic processing, it may also be a requisite of the temporal organisation of events which unfold over time, a process which is vitally important in both domains.

1.2 Evidence for a shared processing of music and speech

The fact that speech and music display acoustic and structural similarities may indicate that the two domains draw upon similar sensory and cognitive resources during perception. This hypothesis has been widely researched, notably by measuring i) the transfer effects of musical training on the processing of speech, and ii) similarities and/or interactions in the online processing of speech and music. Whilst ii) aims to investigate domain-general processing overlaps, i) is able to highlight these overlaps, since musical training incurs specific plastic changes. Evidence from these two lines of research will now be presented. The effect of musical training on speech will be considered first. The functional neuroanatomy of sound processing will be outlined in cohort with this evidence. Second, evidence for the sharing of online processing resources will be considered.

1.2.1 Musical training: Subcortical and Cortical organisation of music and speech processing

Whilst subcortical structures seem to be devoted to encoding more basic acoustic parameters of speech and music, cortical structures may be involved in more complex processing feats which are required to make sense of these sensory inputs, such as the extraction sequential or hierarchical rules from the auditory signal. Studies investigating musical training can assess the long-term impact of training on functional brain activation and neuroanatomy (using functional and structural neuroimaging measures, review Herholz & Zatorre, 2012), as well as on the online responses to music and speech in comparison to non-trained individuals (using electrophysiology or behavioural measures).

1.2.1.1 Subcortical Structures

After reaching the auditory nerve, sound information is relayed to the primary auditory cortex via the cochlear nuclei, superior olivary nuclei, inferior colliculus and medial geniculate body. Efferent fibres also feed back from the cortex, allowing for reciprocal connectivity between the auditory cortex, medial geniculate body and inferior colliculus, thus enabling top-down neural plasticity of subcortical structures (Huffman & Henson, 1990). The presence of these top-down inputs means that intensive musical training can impact upon the early encoding of speech (e.g. Wong et al., 2007). Patel (2011) outlines five conditions (Overlap, Precision, Emotion, Repetition and Attention) which are said to be requisites for music-induced impacts on subcortical speech encoding. First, there should be an anatomical ‘Overlap’ in the structures involved. ‘Precision’ refers to the fact that greater precision (pitch or temporal precision, for instance) and thus, higher demands, is required for music than it is for speech. ‘Emotion’ refers to the fact that musical engagement is emotionally rewarding, thus favouring plasticity. ‘Repetition’ is self-explanatory; repetition is required for brain plasticity to occur. ‘Attention’ refers to the fact that music engages focussed attention, which drives plasticity more than passive engagement. Patel proposes that, when these five conditions are met, music has the capacity to induce subcortical neuroplastic changes which also benefit the encoding of speech.

The effect of musical training on subcortical structures has been investigated by measuring the complex Auditory Brainstem Response (cABR); evoked potentials recorded from the scalp which track the encoding of sound from the auditory nerve to the midbrain. Broadly, cABRs consist of an onset response, and a Frequency Following Response (FFR), both of which have been used to measure the sensory encoding of the transitory and stationary signal, respectively. The FFR reflects the processing of spectral characteristics of sound at the inferior colliculus (Greenberg et al., 1987) and has been used to reveal a more faithful linguistic pitch processing in those who have experience in or training with a pitch-based language (Johnson, Nicol & Kraus, 2005; Song et al., 2008). These findings demonstrate that experience-dependent plasticity can occur in the brainstem, a mechanism which was previously not thought to be possible. This thus raised the question as to whether musical experience can also induce subcortical plasticity, and in such a way that processing of acoustic properties of speech also benefits. In a study by Wong et al. (2007), it was found that musicians’ FFRs display a more robust encoding of linguistic pitch than do non-musicians’ (Figure 1.1), suggesting this to indeed be the case.

As well as pitch, subcortical responses reflect the efficiency with which sound onset information is transmitted. Musacchia, Sams, Skoe & Kraus (2007) have extended findings of Wong et al. (2007), showing that musicians not only show a more faithful encoding of pitch (measured by FFR) but also, during audio-visual conditions (in which videos of the sound being made were viewed), there is a more synchronised neural response to sound onsets in both speech and music (as measured by ABR onset latency) (Figure 1.2). The presence of strong cortical-subcortical interactions additionally means that this plasticity is strongly related to auditory cortex responses - this is perhaps not surprising considering that subcortical plasticity may occur as a result of efferent inputs which run from the cortex (Musacchia, Strait & Kraus, 2008). Musacchia et al. (2007) also suggest that, as well as through efferent 'top-down' inputs, brainstem plasticity could occur through a strengthening of afferent connections which are responsible for encoding sound.

This enhanced neural response to speech sounds is thought to explain musicians' greater ability in distinguishing between phonemes which are acoustically very similar (/ba/, /da/ and /ga/) (Parbery-Clark et al., 2012). The possibility for subcortical plasticity to occur through training thus has clinical implications for populations who display a deficit in temporal processing, proposed to underly speech impairments such as dyslexia (Overy, Nicolson, Fawcett & Clarke, 2003; Overy, 2003) and specific language impairment (SLI) (Corriveau & Goswami, 2009). For instance, Russo et al. (2005) have found that auditory training with dyslexic children does indeed enhance the neural synchrony to speech sounds at the level of the brainstem.

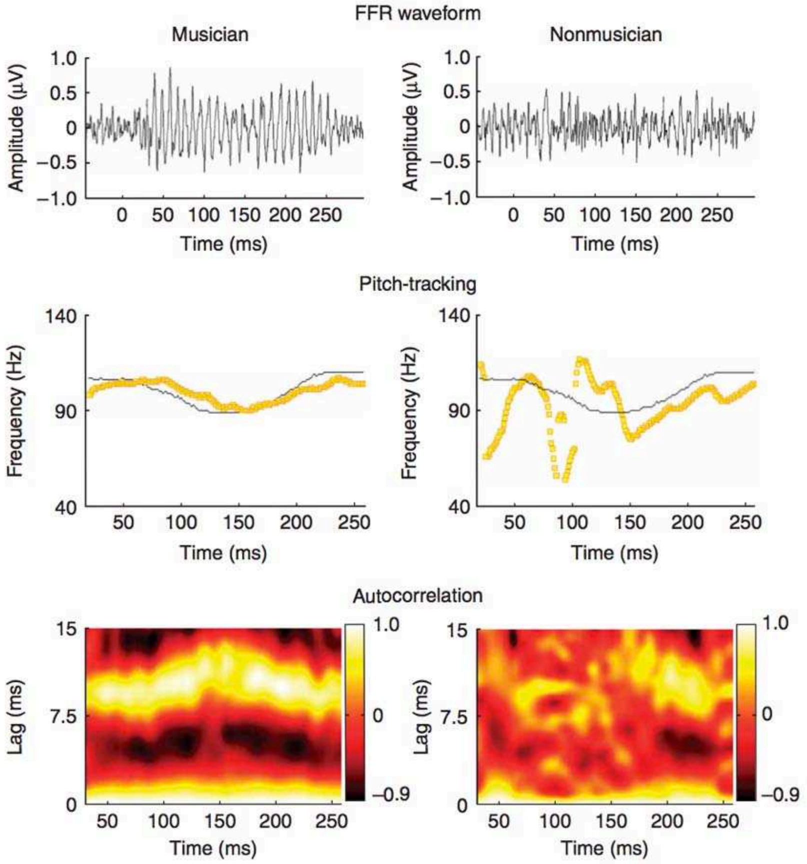


Figure 1.1
Wong et al., (2007): Musicians show a more faithful encoding of linguistic pitch. Top: FFR waveforms elicited by a dipping contour. Middle: trajectories of brainstem pitch tracking (yellow line) to the tone (black line). Bottom: Autocorrelograms of the FFR waveforem: colour indicates the degree of correlation. Musicians' autocorrelogram clearly follows the pitch contour whilst non-musicians' are more diffuse and the highly correlated regions not localised to the frequency of the stimulus.

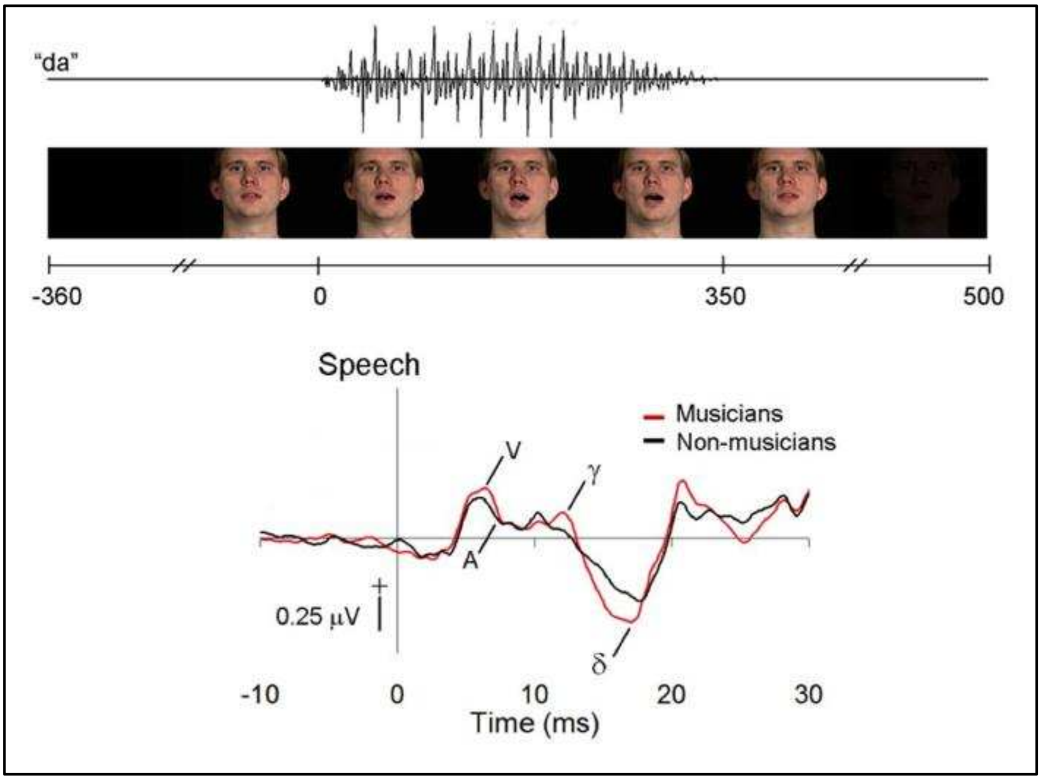


Figure 1.2
Musacchia et al. (2007): Averaged musician and non-musician ABR responses to audiovisual speech (e.g. of stimuli in top panel). Prominant peaks of the onset response are indicated by V, A, γ & δ: δ latencies were earlier in musicians.

Whilst functions of the basal ganglia (BG) and cerebellum (CB) have been primarily linked to motor behaviour, these subcortical structures are also involved in the temporal processing of sound (review Ivry & Spencer, 2004). Indeed, the BG have been proposed to play a more general role not only in in sequential motor planning (Graybiel, 1997) but also sequential processing in general (Brown, 1999). Notably, sequential processing is not only important for motor behaviour, but also for music and speech perception. The BG has been implicated in the temporal perception of both speech (Kotz, Schwartze & Schmidt-Kassow, 2009; Lieberman, 2000), as well as the processing of temporal structures in music, such as beat and interval timing (Buhusi & Meck, 2005; Grahn, 2009; Grahn & Brett, 2007, 2009; Meck & Benson, 2002; Schwartze, Keller, Patel & Kotz, 2011).

Similarly, the CB is recruited during auditory perception and may play a purely sensory role in auditory processing (Petacchi, Laird, Fox & Bower, 2005). The CB is associated with timekeeping/production of both motor and sensory information, and at the order of milliseconds (e.g. Ivry & Keele, 1989), in both speech (Ackermann, Gräber, Hertrich & Daum, 1997) and music (Harrington et al., 2004). Considering that CB regions have been found to be connected directly to the cochlear nuclei (Wang et al., 1999), it has been proposed to provide an interface between auditory and motor information (Huang & Liu, 1990) and to be vitally involved in sensorimotor integration (Baumann et al., 2005). Together with the pre-supplementary motor area (pre-SMA) and supplementary motor area (SMA), the BG and CB are thought to comprise a pacemaker network which provides a template for speech production (additionally involving the thalamus, motor areas, and the frontal and temporal cortices) (Kotz & Schwartze, 2010).

Musical expertise also results in different functional activity of these areas during music production. In particular, musical training is believed to result in a lower processing load and greater efficiency for the production of temporally accurate, sequential motor movements: this can be visualised as a decrease of activation in brain regions associated with music production. During execution of finger movements, violin players show an absent BG activity compared to amateurs (Lotze et al., 2003) and keyboard players show a decreased pre-SMA activation during bimanual movement (Jäncke, Shah & Peters, 2000). Musicians also show an enhanced connectivity between the BG and SMA, which this is thought to reflect a strengthened audio-motor connectivity through training (Grahn & Rowe, 2009). Complementary to findings that musical training induces cerebellar structural changes (Gaser & Schlaug, 2003; Schlaug, 2001), amateur and professional musicians and also show a different localisation of CB activation during music-related motor execution (Lotze et al., 2003) - also possibly related to a more efficient audio-motor integration - and musicians display a reduced CB activity compared to non-musicians (Koeneke, Lutz, Wüstenberg & Jäncke, 2004).

Overall, training-induced functional changes of the subcortical structures (such as the inferior colliculus, BG and CB) can occur through music training. These plastic changes may subsequently impact on the encoding and processing of speech. Whilst plastic changes at the inferior colliculus seem to impact upon the initial, sensory encoding of sound, training-induced changes of the BG and CB could potentially modify a timing network which is engaged by both music and speech, such as that proposed for temporal speech processing by Kotz & Schwartze (2010).

1.2.1.2 Cortical Structures

The primary auditory cortex, housed within Heschl's gyrus, is the first cortical structure to receive auditory information, via the thalamus. It is thought to be involved in processing low-level acoustical features of sound, such as pitch and intensity. Both musical (Gaser & Schlaug, 2003; Menning, Roberts & Pantev, 2000; Pantev et al., 1998; Schneider et al., 2002, 2005) and linguistic (Wong et al., 2008) pitch training induces plastic changes in the auditory cortex. These plastic changes imply altered functional properties, which can be reflected by differences in event-related potentials (ERPs). ERPs are the average of stimulus-time-locked electroencephalography (EEG) recordings.

Early ERPs such as the N100 and P2 are thought to originate from the belt and parabelt regions of the auditory cortex (Liégeois-Chauvel et al., 1994), and the Mismatch Negativity response (MMN) also from the auditory cortex (Näätänen, 1990). The MMN reflects the pre-attentive detection of a violating element (a violation in amplitude, duration, or pitch) within a repetitive sequence of sounds (Näätänen, Paavilainen, Rinne & Alho, 2007) and is thought to be an automatic, short-term neural trace of regularities in the auditory environment (Näätänen & Winkler, 1999). On hearing tones, musically-trained individuals show a greater MMN amplitude (Meyer et al., 2011) and N100 amplitude (Bosnyak, Eaton & Roberts, 2004; Fujioka et al., 2006; Pantev et al., 1998; Schneider et al., 2002; Shahin et al., 2003). This, broadly, marks musicians as having a more efficient processing in auditory cortical areas. Noteworthy here, is that the MMN and N100 may be generated from the same set of neurones (Jääskeläinen et al., 2004).

The planum temporale (PT) is localised posterior to Heschl's gyrus, and in most individuals (about 65% of the population) is most prominent in the left hemisphere (LH) (Becker, 2002). With regards to language, the PT is involved in word recognition, verbal memory and language comprehension; it comprises much of Wernicke's area. It has been said to play an important role in sensorimotor integration in speech production (Hickok, Okada & Serences, 2009), and has been described as a 'computational hub' which is involved in processing spectro-temporal information which is to be relayed to higher cortical areas (Griffiths & Warren, 2002). A more pronounced PT asymmetry has also been associated with absolute pitch abilities (Keenan, Thangaraj, Halpern & Schlaug, 2001).

Passive music listening recruits the LH PT to a greater degree in musicians compared to non-musicians, in particular when training was started at a younger age (Ohnishi et al., 2001), and this is thought to reflect a greater capacity for pitch encoding. This complements structural findings: musicians have larger LH PT (Schlaug, Jäncke, Huang & Steinmetz, 1995; Zatorre et al., 1998). In terms of cross-domain transfer effects, these music-induced structural and functional changes of the PT are proposed to account for a greater verbal memory of musically trained individuals (Franklin et al., 2008; Ho et al., 2003).

The neural bases of processes which occur following sound input to the auditory cortex remain largely unknown. However, both speech and music engage cortical and subcortical networks which simultaneously interact with and modify one another. In the case of speech, Hickok & Poeppel (2007) have proposed a dual-stream model for speech processing. This model describes the simultaneous audio-motor and memory retrieval processes which occur during speech processing: from the auditory cortex, a dorsal stream is proposed to map sensory and phonological representations onto their articulatory motor representations (largely involving LH structures) and a ventral stream maps these representations onto lexical conceptual representations (involving structures bilaterally). A similar model of speech perception has been proposed by Scott & Johnsrude (2003), who identify anterior streams (for mapping phonological cues to lexical representations) and posterior streams (mapping sound to articulatory motor representations) running from the auditory cortex.

The audio-motor aspect of speech processing, as identified by these models, is said to be particularly important during speech acquisition (Hickok & Poeppel, 2007). In music, audio-motor loops are of equal importance (Zatorre, Chen & Penhune, 2007). Evidence even suggests that audio-motor processes in both speech and music may similarly recruit bilateral superior temporal regions of the brain (Dick et al., 2010).

The inferior frontal gyrus (IFG) (which contains Broca's area in the LH) is a key area not only implicated in speech processing (Penfield & Roberts, 1959; Stromswold, Caplan, Alpert & Rauch, 1996), but also in processing hierarchical structures in general, such as in movement, hand and mouth gestures, musical syntax, calculation, working memory, and audio-motor integration (review Fadiga, Craighero & D'Ausilio, 2009; Koelsch et al., 2002; Kotz & Schwartz, 2010; Levitin & Menon, 2003; Maess et al., 2001). Indeed, (bilateral) IFG activation has been associated with the processing of musical information which requires a greater degree of structural (harmonic) integration (Tillman, Janata, & Bharucha, 2003). Musical training may also result in structural changes to the IFG. For instance, musicians' IFG is found to contain more grey matter (bilaterally) than non-musicians (James et al., 2013; Sluming et al., 2002). This could indicate the greater cortical efficiency with which musicians are able to integrate incoming information - be it music, speech, or hierarchical sequences relayed by other modalities.

1.2.1.3 Conclusions: Training-induced plasticity

Overall, musical training has been found to induce subcortical and cortical plasticity of areas not only implicated in music processing, but also speech processing. At the subcortical level, music can induce plasticity of the inferior colliculus (which is involved in initial encoding of sound), the BG (which play a role in regularity extraction, sequential perception and production) and the CB (which has a role in pre-attentive interval timing, and audio-motor integration). The BG and CB may also be key structures in subcortico-cortical networks involved in temporal processing and predictive coding of speech (Kotz & Schwartze, 2010), networks which are potentially also engaged by music.

At the cortical level, training-induced plasticity may impact on speech via auditory cortical areas (which processes spectral and temporal information), the PT (which plays roles in processing spectral information, sensorimotor integration and verbal memory in speech) and the IFG (involved in structural integration). Training-induced plasticity may also alter audio-motor connectivity and integration. Since audio-motor processes in speech and music may have a common neural basis (Dick et al., 2010), training may also impact upon audio-motor processes, through, for instance, plasticity of a posterior stream involved in audio-motor integration in speech (Scott & Johnsrude, 2003).

1.2.2 Online processing of music and speech

Processes dedicated to both speech and music processing can also be investigated by tracking the time course of brain responses. One approach has been to identify the whether online cognitive operations required for various levels of speech processing (e.g. intonation, syntax and semantics), are required for similar, music-based operations (e.g. melody, harmonic relations, meaning). As noted, this 'dissection' of music and speech allows their constituent elements to be compared. Whilst this dissection may be based on characteristics/functions of speech/music (e.g. syntax/harmonic relations), it may be more informative to identify the common sensory and cognitive operations underlying their perception (e.g. structural integration) (see Besson & Schön, 2001). This may sidetrack arguments for a modularity of music processing (Peretz & Coltheart, 2003). Evidence for common underlying operations for the processing of intonation/syntax/semantics and their musical counterparts will be outlined here.

The processing of linguistic intonation contour and musical melody has been found to be enhanced in musically trained individuals (Schön, Magne & Besson, 2004). In general, musically trained (vs. untrained) individuals elicit greater early negative ERPs in response to not only musical pitch violations (e.g. Fujioka et al., 2006) but also linguistic pitch violations (Magne, Schön & Besson, 2006; Marques, Moreno & Besson, 2007; Moreno et al., 2009; Schön et al., 2004). Linguistic pitch violations have also been found to elicit a reduced later positivity in musicians (Moreno & Besson, 2006; Moreno et al., 2009). Taken together, this evidence supports the view that music-induced auditory cortex plasticity results in a more automated processing of linguistic pitch relations - seemingly through a greater efficiency of neural networks which are recruited for the processing of both linguistic intonation and melodic processing.

Syntactic processing in language describes the ability to encode grammatical sequences. An analogy in music has been made, which is the processing of harmonic relations (e.g. Koelsch, 2005; Patel, 1998; Patel et al., 1998; Slevc, Rosenberg & Patel, 2009). In both instances, events (words/chords) should unfold over time in accordance with grammatical 'rules', and should be integrated with a context provided by preceding events. Whilst the cognitive operations themselves are not thought to be shared by the two domains, due to differences in their 'form, use and purpose' (Patel., 1998), both cases are thought to draw upon a process of structural integration. This process may rely on shared neural substrates. Indeed, electrophysiological responses to grammatical violations in both speech and music are similar: syntactic incongruities in both domains elicit a P600 component (Patel et al., 1998) (a component usually associated with speech syntax processing, see Hagoort, Brown & Groothusen, 1993). The P600 has thus been considered to index an online process of structural integration, in general (Patel et al., 1998). These findings led to the Shared Structural Integration Resource hypothesis (SSIRH), which claims that music and speech shared cognitive resources for structural integration (Patel, 1998, 2012).

Support for the SSIRH also comes from Koelsch et al. (2005) who presented words (visually) with simultaneous chords. Sentence-final words either presented a grammatical violation or not, and were accompanied by a chord which either violated the harmonic context, or not. For conditions in which a syntactic incongruity was present in both speech and music, an interaction occurred between two components associated with language and musical violations (the left anterior negativity (LAN) and early right anterior negativity (ERAN), respectively). More precisely, the LAN was reduced in these double-incongruous conditions, presumably because the processes underlying the LAN/ERAN responses were competing for the same resources. Further electrophysiological support comes from the study of language-impaired populations; individuals with syntactic processing deficits in speech (SLI children) do not elicit an ERAN in response to unexpected events in music (Jentschke, Koelsch, Sallat & Friederici, 2008). This shows that their syntactic speech processing deficit extends to music syntax processing and that similar resources may be required for structural integration in both domains.

The SSIRH has also been supported up by behavioural evidence: Fedorenko et al (2009) measured comprehension accuracy of sung stimuli, in which target words (syntactically expected or unexpected) were sung on either expected or unexpected pitches. The authors also measured this effect at varying levels of syntactic complexity. At the linguistic level, unpredictable syntactic conditions were either subject-extracted sentences or object-extracted sentences (the latter being considered to be more syntactically complex), and at the musical level incongruencies were based on short- or long-term harmonic distances. They found that these different levels of linguistic and musical structural complexity interacted, and thus claim that the two domains were competing for neural resources required for structural integration. As noted, one candidate area for these shared neural resources is the IFG (in particular, Brocas Area), an area thought to be involved in structural integration of incoming information, across not only domains but also modalities (see Fadiga et al., 2009; Hagoort, 2006).

One consequence of structural integration might be said to be the extraction of meaning. Both speech and music may be said to have a semantic component, though evidently music not in strictly the same sense as language; music does not convey explicit meanings in the same way that language does. However, music is capable of conveying concepts through associative meaning to external objects or events (Daltrozzo & Schön, 2009; Koelsch et al., 2004), and can also convey a meaning which is self-referent (Steinbeis & Koelsch, 2008). For instance, we implicitly experience pitches in music as adhering more or less to expectations, based on a given pitch context (Krumhansl, 2000). In this way, a given harmonic context gives rise to harmonic expectations (Bharucha & Stoeckig, 1986), and these tension-resolution patterns displayed by music are what have long been said to provide listeners with what we understand as musical meaning (Meyer, 1956). Notably, the N400 is thought to be a marker for semantic integration in speech, and the N500 is thought to index semantic integration in music (Steinbeis & Koelsch, 2008). These ERP components have been considered to be similar in origin, and, again, for the left IFG - an area also associated with semantic speech processing (Bookheimer, 2002) - to be involved in these operations (Koelsch et al., 2004; Levitin & Menon, 2003; Steinbeis & Koelsch, 2008).

For final brief consideration here is that of pragmatics. In language, pragmatics refers to an understanding which is not only dependent on semantics, but also on an understanding of the context within which things are being said, the intentions of a speaker, or a knowledge outside what is actually being said. These interactions are also essential for the communication between musicians about musical intentions in music performance, especially during improvisation. Through 'dialogue' in both domains, the intentions and a 'higher meanings' of communication can be communicated. Timing, in particular, (turn-taking, interruptions, backchannelling) is an important element of both linguistic and musical dialogue (Gill, 2012; Gill, Thompson & Himberg, 2012). The pragmatics of music and speech has been said to present one of the most fruitful cross-domain comparisons, insofar that spontaneous composition is akin in many ways to spontaneous speech (Coventry & Blackwell, 1994). Pragmatics in both domains may also work around similar principles which invoke emotional responses (Coventry & Blackwell, 1994). Indeed, the use of emotional prosody/intonation is important in conveying intentions in both speech and music, and the processing of these cues may draw upon similar neural resources (Thompson et al., 2004).

In his resource-sharing SSIRH framework, Patel (1998, 2012) suggests that online processing resources (for structural integration) may be shared, whilst domain-specific representations are stored independently. As a side-note, this framework is thus able to explain the dissociations found in music and speech processing such as is the case with amusia (Peretz et al., 2002), whereby pitch processing in speech is normal, but musical pitch processing may be impaired (however, see Patel et al., 2008). From the evidence outlined here, it seems that ‘structural integration’ may not only required for syntactic processing, but also the processing of semantic and pragmatic information. Structural integration is, additionally, not only dependent on what speech/music events occur, but also when they occur. There must be, therefore, a vital role of temporal processes. Chapter 2 will present evidence that temporal processes are similarly engaged by both speech and music, and that these processes can impact on both domain-general cognitive and sensory processes.

1.2.3 Conclusions Section 1.2 and Introduction to Chapter 2

A vast amount of research has found that musical training can induce structural and functional changes in the brain. Of particular interest is how this plasticity impacts upon speech processes subsided by these same regions or networks. As well as this neural ‘Overlap’, Patel (2011) outlines four other conditions (Precision, Emotion, Repetition and Attention) which, when met, allow music to induce subcortical neuroplastic changes which also benefit the encoding of speech. These conditions also apply to the music-induced plasticity of cortical structures.

Much of the research described thus far have investigated the influence of musical training on pitch processing aptitudes. Musicians are also timing experts. Since timing is a critical aspect of speech processing, too, musical training may enhance temporal processing abilities in general, i.e. temporal processes which may also be involved in speech processing (see Chapter 2.3).

Temporal processes are also considered to guide online perception, regardless of training. This may be important for various levels of speech and music processing. For speech, temporal processes may impact upon phonemic, word, syntax and lexical-semantic processing. For music, temporal processing may impact on timbre, pitch, melody, and harmony perception, for instance. Notably, whilst the functions of these domain-specific levels may not be shared (e.g. music does not require phoneme perception), there may be domain-general principles which are shared (e.g. both domains require sound analysis, which is important for both phoneme and timbre perception). Temporal processes may impact on different levels within domains via these domain-general processes. Overall, this implies that engagement of temporal processes through music may impact on speech processing, via domain-general principles of cognition (Figure 1.3). Candidates for these shared temporal processes and the extent to which speech and music may engage them will be considered in detail in Chapter 2.3 and in the Discussion section.

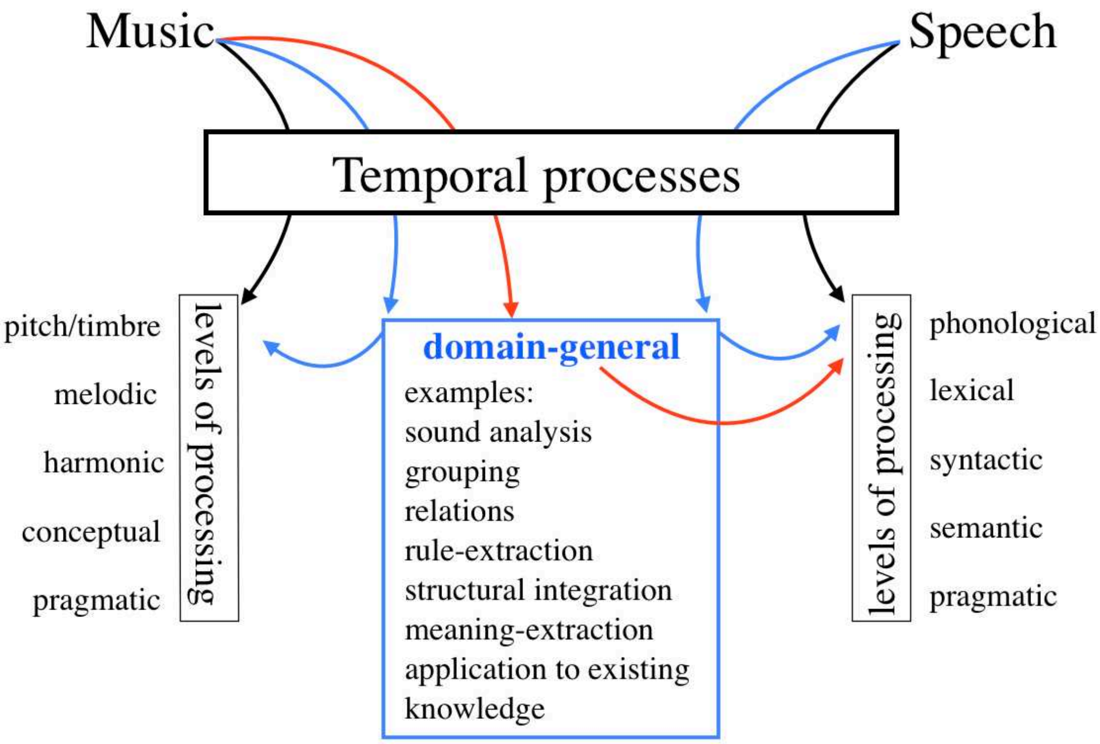


Figure 1.3
Music and speech may engage similar temporal processes. This may impact on different levels of both speech and music processing (black arrows). From a more cognitive standpoint, this may occur via domain-general processes (blue arrows). Evidently, these domain-general processes (examples in blue box) are important in both speech and music. Music may therefore potentially impact on levels of speech processing (as we have recognised them) via domain-general processes which are sensitive to temporal structure (hypothesis) (red arrows).

Chapter 2

TEMPORAL STRUCTURES: Music and Speech

2.1 Temporal structures in music

Music is dependent on a rule-based, temporal unfolding of auditory events. Temporal organisation in music can be described in terms of three interdependent temporal features: beat, metre and durational patterning. Musical 'beat' can be described as the detection of regular, recurrent auditory events which are perceived as isochronous pulses (Cooper & Meyer, 1960). Metre can be defined as an emergent structure which overlays onto beat, resulting in a hierarchical organisation and grouping of auditory events (London, 2004). For instance, Western music most often displays a binary, march-like metre (12 12 12) or a ternary, waltz-like metre (123 123 123). Durational patterning can be described as a third level which is nested within (or can itself infer) both beat and metre.

Whilst these definitions remain somewhat technical, the relationship between the physical aspects of musical temporal structures and their perception is not a 1:1 relationship. For instance, isochrony (beat) is a perceptual phenomena which can be perceived in the absence of true, physical isochrony (Large, 1999; Lehiste, 1977), and metre perception can be modified at will (in the case of subjective rhythmisation, e.g. Iversen, Repp & Patel, 2009). Temporal features of music can therefore be defined by 1) their acoustic, physical characteristics, and 2) the perception of these characteristics. The disparity (or similarity) between the physical and the perceptual aspects allows insights as to how temporal structures are processed in the brain, and can increase our understanding about how a physical stimulus and the brain interact to form a percept.

2.1.1. Physical correlates of temporal structures in music

The beat, metre and durational patterning of musical structure can be measured in terms of both their exact timing and their acoustic features. For instance, interonset intervals (IOIs) can indicate the regularity of a 'beat', as well as the beat tempo. Taking physical measurements of 'beat' may be useful when assessing beat production ability. In these cases, beat synchronisation is assessed by measuring the temporal asynchronies (usually negative, i.e. early) between a produced (e.g. tap) and target (e.g. metronome) event, with standard deviations of these asynochronies taken to infer beat synchronisation ability, for example (also see Farrugia et al., 2012; Repp, 2005; Repp & Su, 2013).

As well as information about 'when' events occur, metre contains information about 'what' will occur. There are acoustic correlates for stressed and unstressed elements in metrical patterns. Stressed events, for instance, display what is described as 'phenomenal accents'; events which show a marked change in duration (e.g. Essens & Povel, 1985; Palmer & Krumhansl, 1990), loudness (e.g. Chen, Zatorre & Penhune, 2006), timbre and pitch (Lerdahl & Jackendoff, 1983, p.17). These features mark events as more or less salient, and in this way are able to infer a metrical structure.

The intentions of musical temporal structures can also be derived from their written notation. In this case, the underlying beat is uninteresting, as it remains isochronous, and the metre is also indicated unambiguously. Durational patterning in this case may be the most interesting thing which can be extracted from notated scores. Durational variability describes the temporal variability of durational patterns. In music, this can be measured by the normalised Pairwise Variability Index (nPVI). Interestingly, analyses of written notation has allowed this method to reveal differences in the durational variability of musical rhythm of different cultures (Huron & Ollen, 2003; Patel & Daniele, 2003; Patel, Iversen & Rosenberg, 2006). Though the nPVI method infers the 'global' variability based only on the variability between adjacent elements, it is a valuable tool in comparative research, not only in assessing differences in durational variability between languages, but also between speech and music (discussed further in Section 2.3).

2.1.2 Perceptual correlates of temporal structures in music

It is thought that beat perception is an innate predisposition of human cognition. Indeed, the ability to perceive both beat (Honing, Ladinig, Winkler & Háden, 2009; Winkler et al., 2009; Zentner & Eerola, 2010) and metre (Hannon & Trehub, 2005) is present from infanthood. As noted, beat itself is largely an endogenously-generated percept and can be extracted from sounds which in reality are not periodic (Lehiste, 1977): for instance, whilst expressive musical performance invariably displays temporal fluctuations, we are nonetheless able to follow the underlying beat. To be sure, beat fluctuations are an important aspect of musical performance and may be paramount in driving our emotional responses to music (Chapin et al., 2010a).

Metre contains not only beat information, but additional hierarchical levels which overlay beat. Similarly, metre is thought to be a universal trait of human cognition (Drake, 1999; London, 2004). It is also an endogenously-generated percept: it has long been noted that the perception of metre can arise even when the stimuli presented are acoustically identical (Bolton, 1894; Fraise, 1982; Woodrow, 1909). It is thought that this subjective accenting of elements occurring within a repetitive sequence is a natural perceptual consequence of the need to 'group' events together, much like the visual gestalt principles (Aksentijevi, Elliot & Barber, 2001; Bregman, 1994; Deutsch, 1999). This grouping phenomenon also results in the formation of metrical expectations (Brochard et al., 2003; Potter, Fenwick, Abecasis & Brochard, 2009).

The availability of different neuroimaging approaches allow us to investigate temporal processing from two different angles. Firstly, electroencephalography (EEG) and magnetoencephalography (MEG) (methods with a high temporal resolution) allow us to follow online changes which occur during the perception of temporal structures. Secondly, functional magnetic resonance imaging (fMRI) and MEG (methods with a high spatial resolution) allow us to investigate candidate brain areas involved in the temporal processing of music. Evidence from these two approaches will now be considered.

2.1.2.1 The online perception of temporal structures and influencing factors

The Dynamic Attending Theory (DAT) posits that temporal structures in music induce endogenous oscillatory activity which is synchronised to the external auditory stimulus (Jones & Boltz, 1989; Large & Jones, 1999). This theoretical framework suggests that temporally regular auditory structures induce synchronised neural activity, whereby neuronal populations resonate at the same frequency as the perceived beat. This neural entrainment to beat can even occur even in the absence of an explicitly presented beat, such as is the case for syncopated rhythms, where a beat is only inferred (Velasco & Large, 2011). Neural resonance to beat is also said to occur in response to auditory information which is, in reality, not isochronous (Large, 2008; Large & Palmer, 2002; Madison & Merker 2002).

The perception of an overlaying metrical structure results in synchronised neural activity which reflects the metrical saliency of the event, and acts to further strengthen beat expectations (Ellis & Jones, 2010; Large, 2008; Large & Jones, 1999; Large & Kolen, 1994; Large & Palmer, 2002; Nozaradan, Peretz, Missal & Mouraux, 2011; Snyder & Large 2005; van Noorden & Moelants, 1999) (see Figure 2.1). Again, this neural synchrony can occur even in the absence of acoustic correlates of metre. For instance, imagined metrical patterns (when isochronous events are subjectively accented) results in a neural entrainment which can be visualised through the amplitude of steady-state EPs: imagined binary structures elicit an increased EEG amplitude at 1/2 the frequency of the beat, and imagined ternary structures elicit an increase at 1/3 and 2/3 of the beat frequency (Nozaradan et al., 2011). Metre perception has also been investigated by measuring oscillatory activity (which measures not only in the time, but also the frequency range), and, similarly to hearing a beat (Fujioka et al., 2009) is associated with synchronised beta-band activity (14 - 30 Hz) (Iversen, Repp & Patel, 2009). This is interpreted to reflect the role of beta band activity in linking endogenous (e.g. a preference for metre perception) and exogenous information (e.g. external auditory stimulus), and, given its role in motor processing (Schnitzler et al., 1997), to indicate a motor influence on metre perception. Neural resonance to metrical structures has also been marked by synchronised gamma band (> 30 Hz) activity (Snyder & Large 2005).

Due to the allocation of attention in time, expectations induced by temporal structures in music are said to impact upon the processing of auditory information. Namely, auditory events which coincide with a point of expectation (e.g. ‘on’ an inferred beat) are said to occur within the attentional ‘window’ for optimal processing. As a result, these events are processed more rapidly than those which occur ‘off’ the beat, presumably due to a reduced processing load at attended points (Barnes & Jones, 2000; Jones, Moynihan, MacKenzie & Puente, 2002; Large & Jones, 1999; Peelle & Davis, 2012; Tierney & Kraus, 2013a; Tillman & Lebrun-Guillard, 2006).

An additional, overlaying metrical structure, or the temporal relations which are implied by metre, are said to strengthen beat expectations and thus further facilitate processing at expected timepoints (Ellis & Jones, 2010; Kung, Tzeng, Hung & Wu, 2011; Large & Jones, 1999). Interestingly, metrical expectations induced by auditory rhythm can also enhance the detection of visual targets (Bolger, Trost & Schön, 2013; Escoffier, Sheng & Sharmer, 2010; Miller, Carlson & McAuley, 2013). This shows that the ability for temporal orienting to enhance processing can span across modalities, and thus it may constitute a general cognitive process.

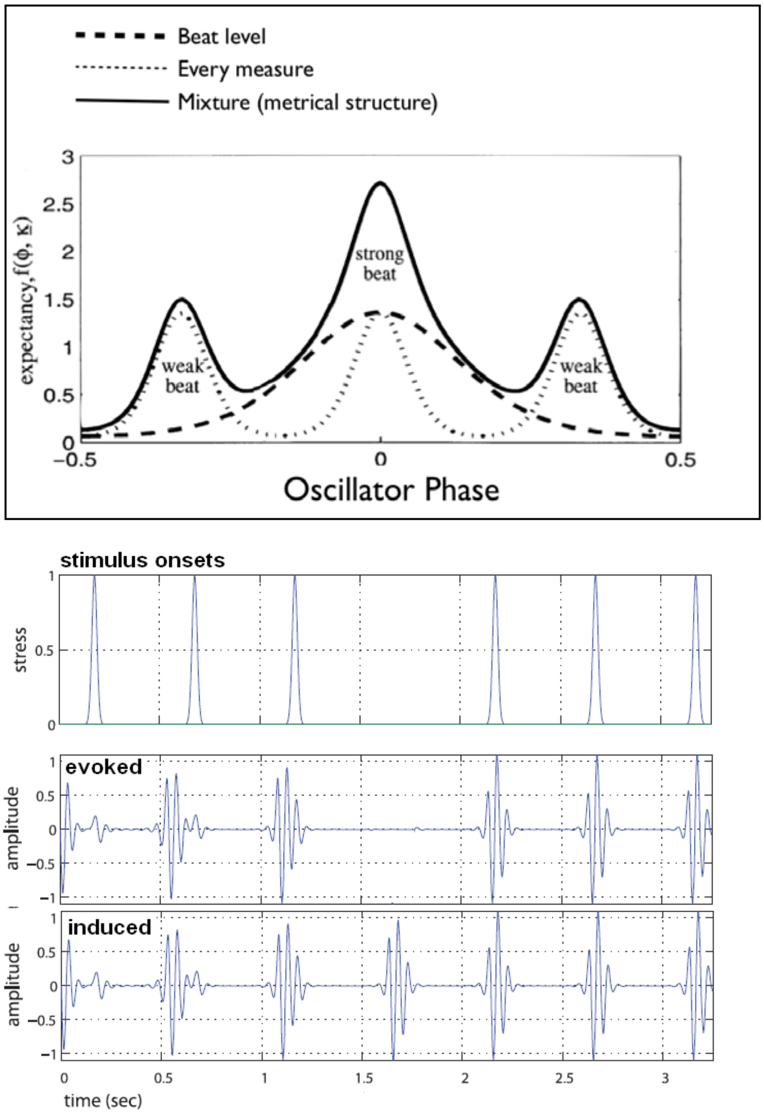


Figure 2.1
Top: Adapted from Large & Palmer (2002). Expectancies for a ternary metre. Multiple oscillators resonate with different metrical levels (beat level and every measure). This activity summates, resulting in a metrical trace (filled line) which represents the temporal expectancies over time. 'Phase 0' simply means on-beat. **Bottom:** Snyder & Large (2005) (adapted): Activity in the gamma band (20 – 60Hz) corresponds to beats in a simple rhythm: induced gamma band activirty persists even when a beat is omitted. This suggests that induced gamma activity may serve as an internal metronome, independent of auditory input.

Expectations induced by a temporal context can be investigated by studying the online response to deviants within a temporal sequence. Ordinarily, ERPs displaying a greater amplitude and later latency indicate that more resources are recruited to process these deviants. Several EEG and MEG studies have reported that temporal deviants induce a MMN/MMNm response. Interestingly, musicians have been found to show a more marked MMN response to metrical incongruities, as well as a more left-lateralised response (as opposed to a more right-lateralised response as is the case with non-musicians) (Vuust et al., 2005, Figure 2.2). Notably, the LH is said to be specialised for fine temporal processing, such as that required by speech (Zatorre & Belin, 2001; Zatorre, Belin & Penhune, 2002), and a similar hemispheric shift occurs for phoneme perception in language learning (Rinne et al., 1999; Shtyrov et al., 2000).

As a side-note, the MMN response might be said to be a pre-attentive equivalent to the N100 response. It has been proposed to be generated by the same set of neurones and falls roughly within the same latency window (150 - 250 ms) (Jääskeläinen et al., 2004). The N100 response is also elicited by both beat and metre violations, and for both musicians and non-musicians (Geiser, Ziegler, Jäncke & Meyer, 2009; Potter et al., 2009; Tierney & Kraus, 2013a). We too have reported a greater N100 response to metrical violations of musical rhythm by speech (Cason & Schön, 2012 - Experiment 1).

A greater amplitude parietal P300 component has also been implicated as a response to deviances in beat (Experiment 1; Lange, 2009) and metre (Vuust et al., 2009). A greater P300 amplitude has also been reported in violations to an imagined metrical structure (Brochard et al., 2003; Potter et al., 2009). Whilst the N100 and/or MMN responses are considered to reflect early, sensory processes (see Brochard et al., 2003), since the P300 is a later response it is considered to reflect a more cognitive (decision-related) or motor (response-related) response to deviant detection, and could reflect the response to the error signal of the MMN (or N100) (Vuust et al., 2009). Since greater amplitude MMN and P300 responses have also been found in response to melodic violations (Trainor, McDonald & Alain, 2002), they may represent general markers for expectancy violations.

Overall, perceived temporal structures in music are thought to induce synchronised neural activity and result in the formation of implicit temporal predictions. When temporal expectations are violated, differences in early (MMN or N100) and later (P300) responses can be observed. Whilst electrophysiology can tell us about the timescale of online processing, it is difficult to localise from which brain regions this activity stems from. Due to its high spatial resolution, fMRI is more able to reliably do so, and will be considered next.

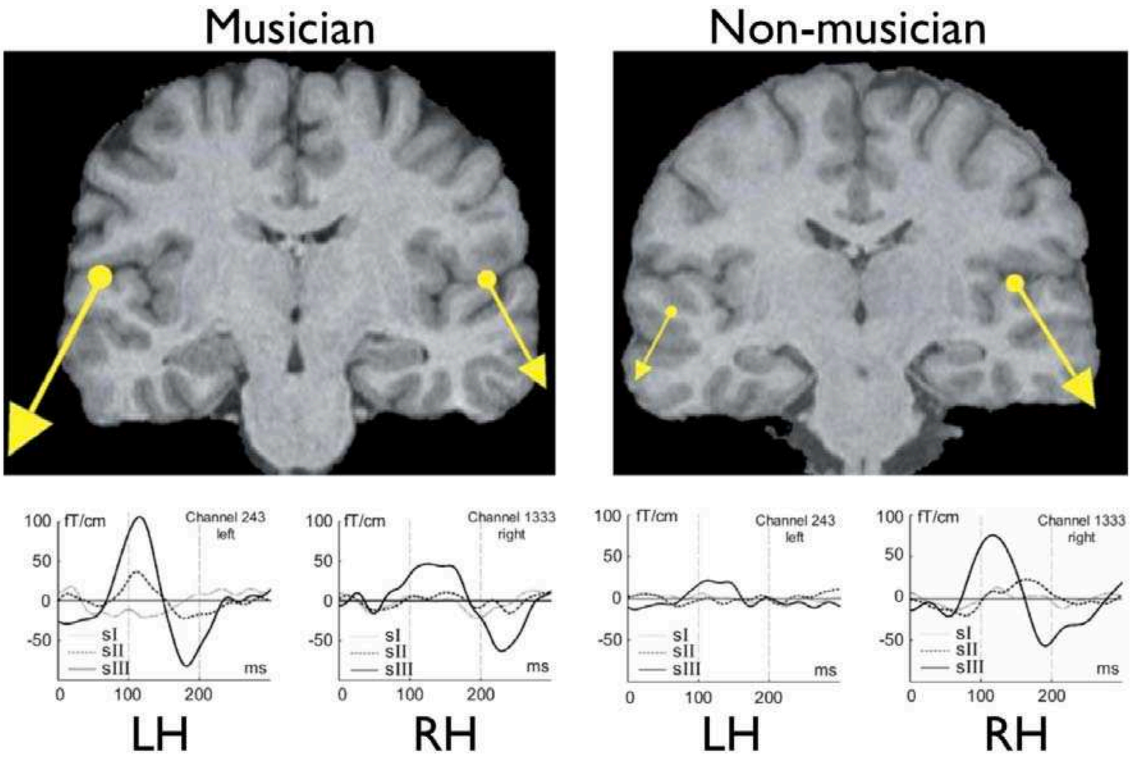


Figure 2.2
Vuust et al. (2005) (adapted).
Musicians show a stronger response to weak metrical deviants, as well as left-lateralised response. **Top:** Musicians show a greater amplitude, and more left-lateralised MMN(m) response to strong metrical deviants ('sIII' conditions) (yellow arrow indicates strength (amplitude) and direction of MMN responses). **Bottom:** MMN responses in the three experimental conditions. sI conditions (grey line) = simple beat (no MMN); sII conditions (dotted line) = syncopated beat (unexpected, but does not interfere with the metrical structure) (only musicians respond to this); sIII conditions (black line) = departure from the beat (greatest violation, MMN elicited by both musicians and non-musicians).

2.1.2.2 Candidate areas for temporal processing

Most of the early work on the localisation of the temporal processing of musical temporal structures has been done with brain damaged patients. Though the processing of musical features (such as pitch and rhythm) must at some point become integrated to form a global percept of music (see Jones et al., 2002; Pitt & Monahan, 1987; Schellenberg, Krysciak & Campbell, 2000; Tillman & Lebrun-Guillaud, 2006), processing of these features may also be differentially impaired, perhaps due to the fact that their integration occurs at later stages of processing (Tillman, 2012; Tillman & Lebrun-Guillaud, 2006). This being said, the perception, attention and memory for pitch relations may themselves be rhythmical (Boltz & Jones, 1986; Jones, 1993; Jones et al., 2002).

Bain damage can result in an impaired pitch processing, whilst temporal processing is spared (Ayotte et al., 2000; Liégeois-Chauvel et al., 1998; Murayama, Kashiwagi, Kashiwagi & Mimura, 2004; Peretz, 1990; Peretz et al., 2002; Peretz & Morais, 1989; Vignolo, 2003), and perception or production of temporal structures can be impaired whilst pitch extraction is spared (Di Pietro, Laganaro, Leemann & Schnider, 2004; Peretz, 1990). This may in part be explained by hemispheric specialisations of the auditory cortex in pitch (RH) and fine temporal (LH) processing (Zatorre et al., 2002; Zatorre & Belin, 2001).

Dissociations are also found to exist within the temporal domain, with distinct brain regions underlying the processing of different temporal features. For instance, lesion studies have revealed dissociations between rhythmic grouping and metrical organisation abilities (Peretz & Morais, 1989), and between metric classification and durational patterning abilities (Wilson, Pressing & Wales, 2002). In particular, lesion studies have revealed the temporal lobes as being key in metre perception (Liégeois-Chauvel et al., 1998), beat production (Fries & Swihart, 1990) and durational pattern reproduction (Di Pietro et al., 2004).

With regards to lateralisation of temporal processes, lesions studies have not provided consistent findings. Beat is said to be generated in the right temporal auditory cortex (Fries & Swihart, 1990; Wilson et al., 2002), consistent with evidence that it is easier to tap the beat with the left hand and overlaying metrical pattern with the right hand (Ibbotson & Moreton, 1981). Whilst this idea of a LH metre processing is supported by functional imaging data, patient studies provide variable evidence; some studies favour the idea of a RH metre processing (Kester et al., 1991; Samson, Ehrlé & Baulac, 2001) but some claim that metre processing recruits both hemispheres (Liégeois-Chauvel et al., 1998).

Variable evidence is also found for the localisation of durational patterning processing: durational patterns may be processed by LH structures (Di Pietro et al., 2004; Polk & Kertesz, 1993; Samson et al., 2001) or perhaps by a wider network encompassing both hemispheres (Schuppert, Münte, Wieringa & Altenmüller, 2000; Shapiro, Grossman & Gardener,, 1981).

In general, lesion studies broadly support the view that temporal processing is not confined to one brain region and that it recruits a widespread activation across both hemispheres (Stewart, von Kriegstein, Warren & Griffiths, 2006). Functional imaging data has supported some of this evidence, and in particular outlines the links between auditory temporal structure and motor behaviour. Such data will be considered next.

Perception of temporal structures is important for both speech and music, and is strongly linked to motor synchronisation (review Grahn, 2012a). Our response to regularly presented auditory events often manifests itself as a spontaneous motor behaviour, such as tapping our foot or swaying our body, for instance (Drake, Penel & Bigand, 2000). The activation of motor regions of the brain during auditory listening of temporal structures may go some way to demonstrate/explain this automatic behaviour.

Synchronisation to auditory rhythm engages the sensorimotor cortex (Rao et al., 1997), the SMA (Lewis et al., 2004), premotor cortex (Chen et al. 2006), and the right CB (Rao et al., 1997; Jäncke et al., 2000). SMA and CB activation is thought to be specific to auditory synchronisation (vs. visual synchronisation) (Jäncke et al., 2000). Interestingly, the activations observed when synchronising to auditory rhythm are also similar in the absence of actual motor behaviour (i.e. listening only).

The activity which is elicited on simply hearing an auditory beat/metre (Bengtsson et al., 2008; Bolger, Coull & Schön, in press; Chen et al., 2008; Chapin et al., 2010b; Grahn, 2009; Grahn & McAuley, 2009), particularly in musicians (Lotze et al., 2003; Meister et al., 2004), involves the BG, CB, pre-SMA and SMA. These activations are thought to reflect the strong audio-motor connectivity between auditory temporal structures and action. This strong audio-motor connectivity is also demonstrated by the activation of auditory areas when subjects perform only rhythmic motor behaviours (silently tapping) (Bangert et al., 2006; Lotze et al., 2003). Similarly, sensory cortices are recruited for both beat listening alone (EEG-based evidence from Lakatos et al, 2008; Snyder & Large, 2005) as well as for beat synchronisation (Rao et al., 1997).

This coupling between temporally organised sound and action can also be strengthened through practise, whereby musical training enhances connectivity between areas involved in beat perception (BG) and motor responses to beat perception (SMA) (Grahn 2009, 2012a). These findings concur with lesions studies implicating the BG to play a key role in beat perception: BG degeneration (in Parkinson's Disease (PD)) results in an impaired ability to make use of beat information (Grahn & Brett, 2009).

Functional imaging data has also supported the idea that there is a LH auditory cortex dominance for the processing metrically regular metrical patterns (vs. complex and non-metrical rhythms) (Grahn & Brett, 2007). The complexity of metrical patterns is also reflected by differences in motor cortex activation (Bengtsson et al., 2008; Grahn & Brett, 2007; Sakai et al., 1999). Metrically complex structures (vs. metrically simple) are also associated with greater prefrontal cortex activity (Bengtsson et al., 2008; Sakai et al., 1999). This may be because metrically complex sequences engage short-term memory processes to a greater degree (prefrontal regions are strongly implicated in working memory processes (Courtney, Ungerleider, Keil & Haxby, 1997)).

This may highlight a noteworthy issue, which is that different processes are required by not only different stimulus features (e.g. greater metrical complexity), but also different experimental tasks. For instance, there may be different neural substrates recruited by memory retention vs. discrimination of metre (Sakai et al., 1999; Wilson et al., 2002), for interval- vs. beat-based timing (Teki, Grube, Kumar & Griffiths, 2011) (which may be engaged to different degrees depending on metrical complexity, see Sakai et al., 1999), and for beat perception vs. beat finding (Grahn & Rowe, 2013).

Finally, whilst the studies outlined so far have attempted to localise beat, metre and durational processing, the way in which temporal information is actually 'used' is also of great interest, particularly in the current context of this thesis. Namely, there is a distinction between perception of temporal structures (described above), and the (implicit) exploitation of this information in optimising information processing (attentional orienting, DAT, Large & Jones, 1999) (Coull, Davranche, Nazarian & Vidal, 2012).

One key brain region implicated in attentional orienting is the LH inferior parietal cortex (IPC). Whilst IPC contribution has been previously shown for visual temporal orienting (Coull & Nobre, 1998, 2008), a more recent study extended this finding to the auditory domain (Bolger et al., in press). In this study, it was found that when metrical sequences (which induced temporal expectations) were followed by a target which occurred at beat locations (i.e. temporally expected locations), the was a greater LH IPC activation. This study also found that connectivity between the LH IPC, sensorimotor cortex, CB, and SMA is stronger when targets are presented at temporally expected (beat) locations.

To briefly conclude this section, functional imaging and lesion data has identified several key brain regions involved in the perception of musical temporal structures (BG, CB, pre-SMA, SMA and prefrontal cortex) and the attentional consequences of temporal perception (LH IPC). Strong audio-motor connections also reflect the spontaneous nature with which we respond to auditory rhythm.

2.1.2.3 The impact of musical training on temporal processing

Electrophysiologiological and behavioural studies demonstrating the effects of musical training on music perception have been outlined in Chapter 1 (Section 1.2.1.) Here, evidence of how musical training impacts upon the perception of temporal structures in music will be expressed in more detail.

Musical training results in a more timely and accurate perception and production of temporal structures (Aschersleben, 2002; Rammsayer & Altenmüller, 2006; Yee et al., 1994; Drake, 1993). Enhanced temporal production of musicians is strongly linked to an enhanced auditory processing, in particular to online (rather than referent) temporal processing (i.e. processes which do not require memory components) (Rammsayer & Altenmüller, 2006). Indeed, online temporal processing differs between musician and non-musicians: musicians display more pronounced and earlier-latency ERPs in response to omitted tones within a sequence (Jongsma et al., 2005; Jongsma, Quiroga & van Rijn, 2004). This indicates that musicians have a more refined temporal pattern (metre) processing and, as a result, are more sensitive to deviants. This lends support to the view that musicians process metrical information hierarchically (rather than sequentially as non-musicians might be more inclined to do) (Drake et al., 2000a; Jongsma, Desain & Honing, 2004; Geiser, Sandmann, Jäncke & Meyer, 2010; Kung et al., 2011). Notably, this hierarchical-based processing is thought to be advantageous in allocating attentional resources (Kung et al., 2011).

Musicians also show a greater MMN response when detecting rhythmic incongruities (Vuust et al., 2005). Vuust et al. (2005) also note that the MMN origin becomes left-lateralised in musicians, presumably to allow for a more efficient processing (since the LH auditory cortex is thought to be specialised for fine-grained temporal processing (Zatorre et al., 2002; Zatorre & Belin, 2001). One consequence of metre perception in both speech and music is the organisation of auditory events into distinct ‘groups’ (rhythmic grouping). Rhythmic grouping has been found to differ between musicians and non-musicians, as indiced by the MMN response to variations in group length (van Zuijen et al., 2004).

To consider, finally, durational patterning, there is evidence that both musicians and non-musicians may draw upon these cues in a similar manner. Both are able to successfully classify the language-origin of music based on durational variabilities of musical elements alone (Hannon, 2009). This may be due to the fact that durational patterning draws more upon memory resources - as noted, musicians do not necessarily show an advantage in memory-based temporal processes (Rammsayer & Altenmüller, 2006). Passive exposure to a particular music culture, however, may influence the way in which music is perceived, particularly in terms of temporal structure (Bigand & Poulin-Charronnat, 2006; Honing & Ladinig, 2009). Saying this, the processing of durational variability seems also to be independent of exposure: non-western non-musicians with little exposure to Western music, are capable of identifying historical distance between different music excerpts of Western music based on differences in their durational variability (Dalla Bella & Peretz, 2005).

To summarise, the perception of beat, metre and durational patterning is present in all individuals, but may be enhanced through exposure or musical training. However, other temporal tasks such as those based on durational patterning or duration judgements (of two tones for example, Rammsayer & Altenmüller, 2006)) do not show an advantage in musically trained individuals, perhaps as they involve a referent memory component which is not particularly important in musical ability.

2.1.2.4 Perception and Production of temporal structures in music

As a final consideration is how the perception of temporal structures in music may impact upon their production. As noted, integration of different musical elements such as pitch and rhythm may occur during later stages of processing. Findings that pitch and rhythm can be differentially impaired suggests that different neural substrates are responsible for their encoding, at least during these initial stages. This is exemplified in the case of amusia, whereby processing of musical pitch is impaired, whilst rhythmic abilities remain intact (Murayama et al., 2004; Peretz et al., 2002). However, it has also been found that whilst the perception of pitch structures may be impaired, production can remain intact (Loui, Guenther, Mathys & Schlaug 2008). This additionally raises questions as to the interactions between perception and production.

In the case of rhythm, beat-deafness is a recently-described phenomenon which is characterised by an inability to synchronise motor behaviour to auditory temporal structures, and is accompanied by an impaired perception of temporal structures (Phillips-Silver et al., 2011). This has thus been interpreted as a disorder stemming from an impaired temporal perception. However, subsequent experiments which investigated this phenomenon in ten times as many subjects, found evidence that impaired beat synchronisation may instead occur through an abnormal mapping of beat perception to action (Sowiński & Dalla Bella, 2013). The authors found that, in some subjects, whilst there was an inability to synchronise to beat or to music, rhythm perception abilities seemed intact. Overall, it seems that whilst temporal perception and production may vitally contribute to one another, this does not necessarily mean that an impairment in production stems from an impairment in perception. This proposition may also be worth considering in the context of speech and language disorders which have been proposed to stem from impaired temporal processing. The links between perception and action, and speech and music, will be considered further in Section 2.3.

2.1.3 Conclusions: Section 2.1

This Section has considered the temporal structures (beat, metre and durational patterning) of music. In particular, investigating the perceptual interpretation of physical correlates of temporal structures allows an insight into the contributions of top-down processes (e.g. subjective rhythmisation) as well as factors which may influence the perception of temporal structures (e.g. musical training).

Overall, the online perception of temporal structures in music seems to induce both synchronised neural activity and synchronised motor activity. This may form the basis of temporal expectation; predictions which can be violated or met, and which can thus influence the processing of auditory events. In particular, the IPC is an area implicated in this temporal orienting effect. Temporal perception also engages motor regions of the brain, suggestive of strong audio-motor links, and can be influenced by long-term auditory experience such as musical training. It may also be influenced by short-term auditory experience such as exposure or priming. Arguably, these points all apply to the perception of temporal structures speech.

2.2 Temporal Structures in Speech

Several linguistic levels contribute to spoken language: phonology, morphology (lexical level), syntax, semantics and pragmatics. Each of these linguistic levels may draw upon different temporal processes. As a first note, however, it is perhaps important to consider that different levels of linguistic processing may interact; this will be outlined here in order to underline the complex nature of speech processing.

Linguistic levels of speech processing may feed into and interact with one another (e.g. Friederici, 2002; Friederici & Kotz, 2003), though it is still unclear as to the nature of these interactions. For instance, based on electrophysiological and brain imaging data, Friederici (2002) proposes a neurocognitive model of sentence comprehension (Figure 2.3). This model outlines both the time course of speech processing, electrophysiological markers and brain regions potentially recruited. Dual stream models of speech processing (Hickok & Poeppel, 2000, 2007; Scott & Johnsrude, 2003) also propose the routes by which sensory (phonetic) information is mapped onto both lexical representations and motor articulatory representations via distinct, simultaneous streams.

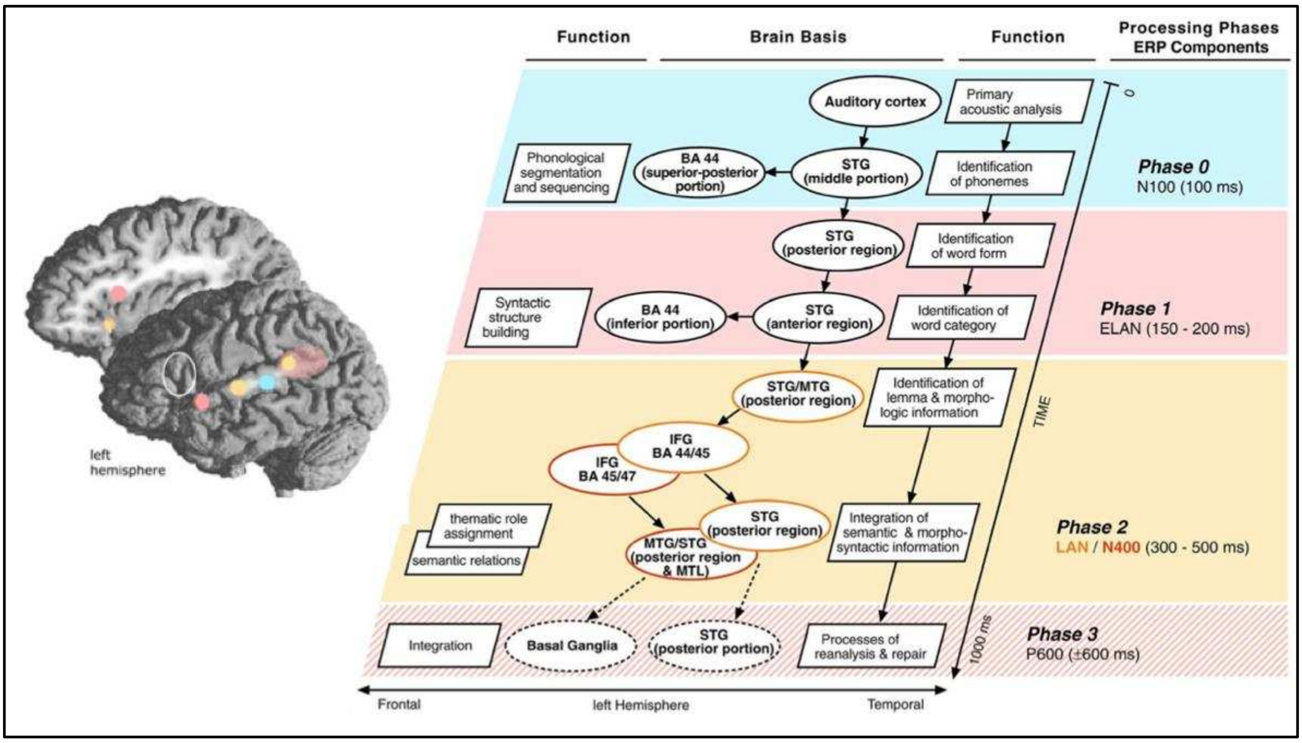
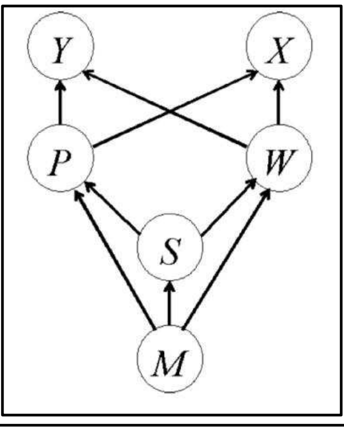


Figure 2.3
Friederici & Kotz's (2003) neurocognitive model of sentence comprehension (adapted from Friederici, 2002). Timecourse, functions, electrophysiological markers and temporal/frontal regions implicated at different stages of processing. Boxes (right) are colour-coded to co-ordinate with dots (left).

Considering that these linguistic interactions exist, temporal structures exhibited by each of these levels (or indeed, the sensitivity to temporal information) may necessarily influence other linguistic levels of processing. In this case, it can be difficult to tease apart the contributions of temporal structure to different levels of processing. To (broadly) exemplify the inter-reliance of these linguistic levels, and in terms of the contributions of 'time': the temporal envelope allows phonemes to be recognised, which in turn allows words to be identified (Shannon et al., 1995). These word forms rely on the ordering of phonemes in time. Syntax depends on a correct temporal unfolding of these words, which can be integrated with semantic representations to form meaning, or a meaning which requires knowledge of the outside world (pragmatics). Dialogue (considered to represent one aspect of pragmatics) additionally depends on a finely-tuned synchronisation with others in the form of cyclic turn-taking and timely interjections (e.g. Bull, 1994). In this case, each linguistic level is dependent on the previous. Whilst this example may infer a bottom-up mode of speech processing, however, this is certainly not the case.

More related to the current focus is to consider the role of suprasegmental properties at each of these levels of processing. The suprasegmental properties of speech may alternatively be described as 'prosody' (pauses, syllable length, pitch intonation and the structure of salient and less salient events, see Zellner, 1994). These features of speech not only contribute to different levels of speech processing (Cutler, Dahan & van Donselaar, 1997; Steinhauer, Alter & Friederici, 1999), but may themselves be governed by higher linguistic levels (Chen & Hasegawa-Johnson, 2004, Figure 2.4). For example, syllabic stress can differentiate meaning between two otherwise identical words, such as affect (A-ffect vs. a-FFECT) (or to put it another way, word meaning governs word prosody).

Figure 2.4
Chen & Hasegawa-Johnson (2004): Their proposed network representing the relationships between different features of an utterance. Arrows represent the (top-down) influence. (Acoustic-phonetic features (X), acoustic-prosodic features (Y), word sequence (word prosody) (W), prosody sequence (of words) (P), syntax sequence (S) and meaning (M) of an utterance).



As well as a lexical function, prosody is said to have an emotional function, a modality function (emphasising, questioning or commanding), and a structural function which is important for parsing and hierarchical organisation. Emphatic prosodic stress, for instance, marks the focus of a sentence in relation to a specific question (e.g. Q: What did he do on Sunday? A: He ran on Sunday vs. Q: What day did he run on? A: He ran on Sunday). Overall, more natural prosody is more easily understood (Hahn, 1999), perhaps as it facilitates lexical-semantic processes (Astésano, Besson & Alter, 2004).

Whilst speech 'prosody' contains several cues (pitch, duration and amplitude), there may be a dissociation in processing its spectral and temporal features (Lancker & Sidtis, 1992). Here, the current focus is more specifically with the temporal features of speech (i.e. syllable saliency and metrical organisation) rather than more spectral-based features (e.g. phrasal intonation).

In terms of neural correlates, temporal processing of speech has been said to recruit the auditory cortex in a scale-dependent manner, one which indicates there to be hemispheric specialisations for temporal and spectral processing. The LH seems to be recruited during tasks requiring a high level of temporal resolution (information occurring within a 20 - 50 millisecond window), whilst the RH preferentially processes information occurring within longer temporal windows (information occurring within a 150 - 250 millisecond window) (Belin et al., 1998; Nicholls, 1996; Poeppel, 2003; Zatorre & Belin, 2001).

In the sections which follow, a focus will be made on those temporal structures which are most analogous to those already described for music: 'beat' will refer to the 'isochronous' aspects of speech which are marked by a stressed syllable (commonly called 'rhythm' in linguistic terminology), 'metre' will refer to the patterning of stressed and unstressed syllables, and durational patterning will refer to the durational patterning between syllable (vowel) onsets. Before considering the parallels between temporal structures in speech and music (Chapter 3), an overview of the temporal structures in speech will now be made - again in terms of both physical correlates and their perceptual counterparts, which can tell us about the relationship between sound and the brain.

2.2.1. Physical correlates of temporal structures in speech

One most obvious example in which the physical correlates of temporal structure has extensively measured is in cross-linguistic research. Correlates for stressed events in speech can be derived from acoustic analyses of duration, pitch and amplitude changes of a syllable, relative to surrounding syllables. Even at this level of analysis, differences between languages become apparent. For instance, primary syllabic stress in French is characterised by syllable lengthening (Dellatre, 1966; Rossi, 1980; Vaissière, 1991), whilst in languages such as English or Dutch, it is also of a higher intensity and can incur a pitch shift (Sluijter & van Heuven, 1996). With differences in the acoustic correlates of stress aside, metrical patterns are realised through stress contrasts in most languages. Though the size of a basic unit in speech has been debated (e.g. phoneme vs. syllable) (Goldinger & Azuma, 2003; Nathan, 2011; Nearey, 1997; Norris & Cutler, 1988; Pitt & Samuel, 1990b; Savin & Bever, 1970), another 'universal' trait of spoken languages seems to be that they are composed of syllable or syllable-like units which are comparable in terms of their rates of unfolding during speech (Greenberg, Carvey, Hitchcock & Chang, 2003).

In attempts to characterise these so-called 'universals' of language, cross-linguistic study has exposed some interesting differences in temporal organisation of different languages. One issue, and one which has been the root of exhaustive debates, concerns cross-linguistic differences in durational and beat aspects of speech. Some first observations were made by Pike (1945) who used the terms 'stress-timed' and 'syllable-timed' to describe the observations that stressed syllables seem to occur at regular intervals in English, whilst syllables in Spanish seem to exhibit an equal duration between adjacent syllables. This distinction between Germanic ('stress-timed') and Romance ('syllable-timed') languages was supported by Abercrombie (1967), who further added to these distinctions 'mora-timed' languages, such as Japanese (a long syllable with lengthened vowels is said to have two moras, whilst a shorter vowel is said to have one mora (Ladefoged, 1975; Port, Dalby & O'Dell, 1987)). Furthermore, Abercrombie (1967) claimed that these distinctions were categorical. Whilst evidence for isochrony in so-called stress-timed speech has not been found (Dauer, 1983; Roach, 1982), several researchers have suggested that languages can instead be characterised according to their durational variability.

The Pairwise Variability Index (PVI) is a metric which has been used to quantify the durational characteristics of language (Grabe & Low, 2002). It calculates the average difference between adjacent vowels or consonants of a given utterance, giving an indication of how variable syllabic intervals are. The normalised Pairwise Variability Index (nPVI) does the same but accounts for changes in speech rate, and has therefore mainly been used for vocalic intervals (which are more sensitive to speech rate).

Languages have been said to fall into distinct categories: ‘stress-timed’ languages are said to have a higher durational variability than ‘syllable-timed’ languages due to their more variable consonant onsets (Ramus, Nespor & Mehler, 1999). However, a subsequent study investigated durational variability using the nPVI, using interval measurements from the vowel onset (Grabe & Low, 2002); the vowel onset is considered to mark the salient perceptual centre of speech (p-centre) - the point at which the energy of the syllable is maximal (Allen, 1972; Kochanski & Orphanidou, 2008; Morton, Marcus & Frankish, 1976; Port, 2003). Enlarging the corpus from 8 to 18 languages, Grabe & Low (2002) found that whilst durational variabilities did indeed differ between languages, they could not be categorised into rhythmic classes. Ramus (2002a), however, re-analysed data from Ramus, Nespor & Mehler (1999) using the nPVI values of vocalic intervals in 20 languages, and maintained that clear clusters of language nPVI values indicate that durational variability can categorise languages. Needless to say, the categorisation of languages into rhythmic classes has been met with some degree of caution (Cummins, 2002; Dauer, 1983; Gibbon & Gut, 2001; Grabe & Low, 2002; Roach, 1982): different languages may show both stress and syllable-timed features (Dauer, 1983; Roach, 1982; Grabe & Low, 2002; Dimitrova, 1998; Mairano & Romano, 2008; Nespor, 1990) and may rather differ along a continuum according to their historical and geographical distances.

Despite their conflicting opinions, these studies have highlighted some methodological issues associated with measuring durational variability in speech with the nPVI, which importantly can account for alterations in speech rate. Most interestingly in the context of this thesis, the nPVI metric can be used not only to investigate across languages, but also across domains (music vs. speech). This will be discussed further in Chapter 3.

2.2.2 Perceptual correlates of temporal structures in speech

Whilst speech does not display the same degree of regularity as does music, there is some evidence to say that humans have a tendency to perceive regularity in speech. In particular, stressed syllables may be perceived as occurring isochronously, forming the 'beat' of speech. In English, it has been found that listeners perceive the onsets of stressed syllables to be more regular than they are in reality (Darwin & Donavan, 1980), supporting Lehiste's view that we extract isochrony from speech, even in the absence of true, physical isochrony (Lehiste,1977).

Specifically, the 'p-centre' is the point to which we assign the 'beat' (Allen, 1972; Port, 2003; Scott, 1998), and to which neuronal oscillators may entrain to (Barbosa, Arantes, Meirles & Viera, 2005). To confound the issue, however, listeners are sensitive to speech timing changes (variations in vowel duration) to the order of milliseconds (Nooteboom, 1972; Schmidt-Kassow & Kotz, 2009b). Perceiving isochrony in speech, therefore, does not arise through an inability to detect small temporal changes, but through a perceptual preference for regularity which results in a temporal categorisation of elements. When temporal changes do not benefit the perception of the speech rhythm (i.e. when they do contribute to the perception of isochrony or regularity), they are probably perceived as violations of the speech temporal structure.

Cummins (2011) addresses the issue of the non-periodic nature of speech and suggests how synchronisation to it may still be possible; whilst speech may not display isochrony, this is not to say it is temporally random. For instance, the fact that speakers are able to synchronise their speech with one another indicates that isochrony is not required for a person to entrain to the speech of others, and that there must be sufficient temporal predictability present in the speech signal for this to be possible (Cummins, 2003, 2009b). Interestingly, the way in which speakers synchronise with one another may also differ according to native language (English vs. Mandarin) (Cummins, Lee & Yang, 2013).

A preference for isochrony has been found in speech production. Speech cycling, the repetition of a phrase, results in a phenomenon termed the 'harmonic timing effect', whereby stressed syllables tend to coincide with fractions of the cycle at the order of 2 or 3 (Port, Tajima & Cummins, 1996; Cummins & Port, 1998; Tajima, 1998). This phenomenon is also thought to reflect speakers' organisation of metrical speech patterns (preferentially binary or ternary, as in music). Similarly, read lists result in a more regular speech timing (Cummins, 2003, 2013). Also notable is that speakers have been found to align their p-centres in turn-taking (Bull, 1984).

This has led to the view that, similarly to neural resonance theories for music perception (Jones & Boltz, 1989; Large & Jones, 1999; review Zanto, Snyder & Large, 2006; Lakatos et al., 2008), the phenomenon of 'periodic' speech reflects the presence of neural oscillations which generate periodic pulses. These oscillations are said to attract prominent motor activity (such as taps to speech (as in Allen, 1972), the production of p-centres in speech, or speech gestures) (Barbosa et al., 2005; Cummins, 2009b; Port, 2003) and to facilitate the perception of speech in more difficult listening environments (review Zion Golumbic, Poeppel & Schroeder, 2012). It has also been proposed that, somewhat similarly to the DAT in music (Jones, 2009; Jones & Boltz, 1989; Large & Jones, 1999), speech engages mutual oscillators for both isochronous events as well as for syllable-sized units (stressed or unstressed) in both speech perception (Arantes & Barbosa, 2010; Giraud & Poeppel, 2012) and speech production (Barbosa, 2002, 2007). In other words, neural oscillations are thought to be attuned to salient and less salient syllables - the metrical structure of speech - and to interact with one another so that there is an online mutual tuning of their cycles. This 'phase resetting' means that neural oscillations can adapt in an online manner to the temporal changes in speech (Peelle & Davis, 2012). This thus accounts for our ability to attend to an incoming speech stream in conditions in which there is other competing auditory input (e.g. cocktail party effect) (Zion Golumbic et al., 2012), as well as to synchronise with speech despite the fact it does not display absolute temporal regularity. This may also explain our preference for extracting temporal regularity in speech perception and may that suggest that speech which is more regularly-produced facilitates perception.

To conclude, though searches for physical isochrony in natural speech has failed, we are able to perceive 'beat' in speech, and in certain speaking situations also show a preference for producing beat-based speech. It may be that there is a processing advantage in perceiving regular, expected, or predictable temporal structures. Whilst this predictability may arise through temporal regularity (e.g. in the case of a repetitive metrical structure), it may additionally arise from a listeners' implicit knowledge about the temporal characteristics of their native language. For instance, French listeners may consistently expect a final-syllable saliency, and may rely less on durational variabilities between p-centres compared to English native speakers (since durational variability is less in French than in English). The potentially advantages to temporal predictability in speech will now be discussed.

2.2.2.1 The developmental function of temporal structures in speech

The speed with which infants acquire spoken language is quite remarkable. Important milestones in the first year of life represent language-independent, universal stages of development (Kuhl, 2004). From birth, infants are sensitive to phonetic contrasts, and at around 3 months they begin to produce vowel-like sounds (Kuhl, 2004). The accurate perception of phonological contrasts is also important for the development of other language-related skills in later development, such as reading (Corriveau, Pasquini & Goswami, 2007).

Before speech perception and production can be achieved, however, infants face the problem of segmenting seemingly continuous strings of sound into individual word units. Whilst statistical probabilities of the language is an important contributing factor, metrical structures in speech also serve as important cues in speech segmentation (Cutler & Butterfield, 1992; Jusczyk, Houston & Newsome, 1999; Kuhl, 2004; Morgan & Saffran, 1995). Whilst newborns and infants seem already to be sensitive to the durational (Nazzi, Bertocini & Mehler, 1998; Nazzi & Ramus, 2003; Mehler & Christophe, 1995; Ramus, 2002b), metrical (Jusczyk, Cutler & Redanz, 1993) and isochronous (Fowler, Smith & Tassinary, 1986) aspects of spoken language, it is thought that the use of these cues for speech segmentation appears between 6 - 9 months of age (Kuhl, 2004; Morgan & Saffran, 1995). The ability to segment the speech stream via stress cues may also aid syntax acquisition. For instance, motherese, which accentuates the metrical aspects of speech, has been found to facilitate the recognition of clause boundaries in 7 - 10 month old infants (Hirsh-Pasek et al., 1987; Nelson et al., 1989).

These early representations formed about the metrical properties of speech can be specific to ones native language and their representation persists into adulthood (Cutler, McQueen, Norris & Somejuan, 2001; di Cristo, 1998; Weber, Hahne, Friedrich & Friederici, 2004). For instance, metrical ‘rules’ which exist for a native speaker of one language are used in the interpretation and segmentation of non-native language contexts, even when these same rules do not apply (Cutler, Mehler, Norris & Segui, 1986, 1992; Otake, Hatano, Cutler & Mehler, 1993). The language-specific use of metrical cues in French, for example, has been used to explain difficulties in encoding foreign stress contrasts which exist in Spanish (Dupoux et al., 1997, 2008) and in German (Schmidt-Kassow, Rothermich, Schwartze & Kotz, 2011). Though French is considered to be a language in which stress cues are relatively unimportant (Ramus, 2002; Dupoux, 1997, 2008; Cooper, Cutler & Wales, 2002), the fact that French is rather metrically consistent language may result in a strengthened representation of speech metre compared to speakers of other languages in which stress position is more variable. In French, stresses occur in consistently word-final and sentence-final locations, with also a word-initial pitch accent (Delattre, 1938; Di Cristo, 1998; Vaissière, 1974). French speakers may therefore have a strengthened representation of speech metre compared to speakers of other languages in which stress position is more variable.

These metrical expectations have also been measured using electrophysiology, where the absence of a P600 component (an ERP elicited by violations of rule-based expectations) in response to stress violations in German implies that French speakers do not perceive the stress pattern present in this language (Schmidt-Kassow et al., 2011). To conclude, the use of metrical cues in early speech segmentation is a universal trait in speech acquisition, whilst the specific metrical strategies in decoding language are language-specific and continue into adulthood.

2.2.2.2 The linguistic function of temporal structures in speech

Once we have formed the temporal 'rules' of our native language, they become interwoven with our processing of phonological, syntactic and semantic features of spoken language. The processing of temporal structures thus also contributes to speech comprehension. The sensitivity of these linguistic levels to temporal features of speech is most commonly investigated by either conforming to or violating temporal expectations.

Phonological processing is considered to form the building blocks of speech perception. In a behavioural study, Quené & Port (2005) investigated the effect of speech regularity on phonological processing. They found that inducing expectations about the beat of speech (where the stressed syllable occurred in a list of words) impacted upon the identification of target phonemes.

This work provided support for the Attentional Bounce Hypothesis (ABH) (Pitt & Samuel, 1990a), which, based on behavioural evidence from a series of experiments, proposes that attention is allocated to coincide with syllables which are expected to be stressed. This work (Pitt & Samuel, 1990a) was originally based on findings that stressed syllables receive more attention and thus contribute more to spoken word perception than unstressed syllables (Cole & Jakimik, 1980; Cutler & Foss, 1977; Shields, McHugh & Martin, 1974), evidence for which continues to mount (Cutler & van Donselaar, 2001; Quené & Koster, 1998; van Leyden & van Heuven, 1996). It aimed to measure how attention is allocated when hearing continuous speech, with the view that temporal regularity in speech may provide a stable structure which a listener can capitalise on to predict the future occurrence of speech events.

This heightened attention to salient or expected events results in an enhanced capacity for phonological processing (also see Shields et al., 1974), and may also be important for other functions such as word segmentation; lexically-ambiguous word sequences which are perceived as being isochronous are also more successfully segmented (Dilley, Mattys & Vinke, 2010; Dilley & McAuley, 2008).

Further research has suggested that attentional orienting in speech occurs through a mechanism of coupled oscillators (Arantes & Barbosa, 2010; Giraud & Poeppel, 2012; Port, 2003; Port, Cummins & Gasser, 1995), an account which is analogous to neural resonance theories in the music domain, such as DAT (Jones & Boltz, 1989; Large & Jones, 1999; review Zanto et al., 2006). Similarly to the perceptual consequences of attentional orienting in speech (Pitt & Samuel, 1990a; Quené & Port, 2005), DAT describes a heightening of processing efficiency at attended timepoints following the entrainment of attentional rhythms (e.g. Jones et al., 2002).

As noted by Pitt & Samuel (1990a), however, speech metre may be less predictable in the context of spontaneously spoken sentences (vs. word lists) and so this attentional effect may contribute less in natural speech. It has even been suggested that we may instead allocate attention to unpredictable points in speech in order to facilitate processing (Astheimer & Sanders, 2011). Nonetheless, the inclination for neural oscillations to undergo phase resetting and period adaption are said to account for these temporal variations (see Figure 2.5)

Whilst beat regularity enhanced phoneme detection, metre expectations (inducing expectations about whether the target word would be iambic or trochaic) did not impact on phonological processing in Quené & Port's study (2005). Nonetheless, behavioural studies have revealed that providing a predictable metrical word pattern can enhance lexical access of visually-presented (Cooper et al., 2002) and spoken words (Dilley & McAuley, 2008; Grosjean & Gee, 1987), semantic integration (Ladefoged, 1975), syntactic processing (Beach, 1991; Eckstein & Friederici, 2005, 2006; Steinhauer et al., 1999) as well as enhancing other functions such as word recall (Gow & Gordon, 1993; Reeves, Schmauder & Morris, 2000). Predictability of speech metre can also come about through familiarity with native lanaguage, which can contribute to lexical access (e.g. Hahn, 1999).

ERP studies, which can measure the online-processing of speech, have also revealed that speech metre contributes to/interacts with syntactic processing (Schmidt-Kassow & Kotz, 2009a; Roncaglia-Denissen, Schmidt-Kassow & Kotz, 2013) and semantic processing (Astésano et al., 2004; Magne et al., 2007; Marie, Magne & Besson, 2011; Rothermich, Schmidt-Kassow & Kotz, 2012). Metrical irregularities in speech have been indexed by a

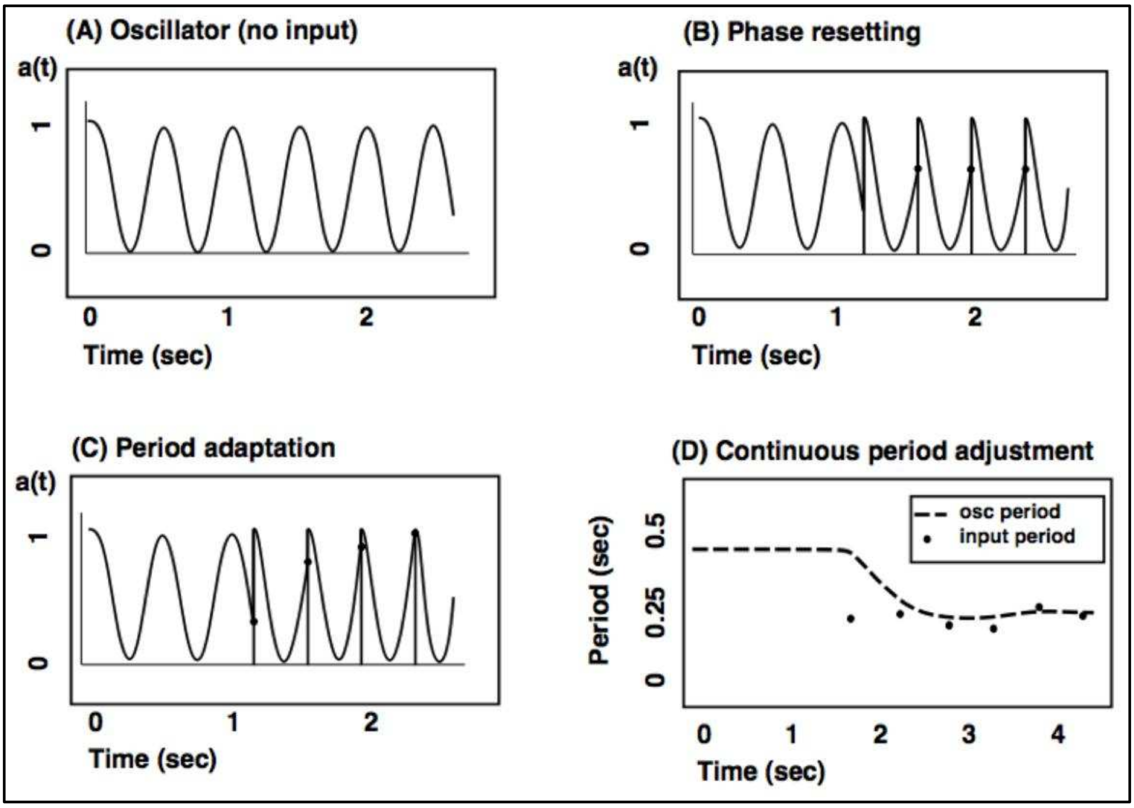


Figure 2.5
Port et al. (2005): Oscillatory adaptations to temporal input. (A) shows the function of an oscillator without an external input. (B) shows an input of periodic pulses (dots) to which neural oscillations adapt to. (C) an oscillator also adjusts its period to become more similar to the external input. (D) shows the oscillator periods (dashed lines) and input periods (dots): especially important in the case of speech, variability in the input pulses tends to be ignored/smoothed out in the oscillators running estimate of the input period.

greater early negativity (Bohn, Knaus, Wiese & Domahs, 2013; Magne et al., 2007; Rothermich, Schmidt-Kassow, Schwartze & Kotz, 2010) which is thought to mark the brain's sensitivity to deviations in metrical/sequential structure, and late positivity (Astésano et al., 2004; Bohn et al., 2013; Magne et al., 2007; Schmidt-Kassow & Kotz, 2009a, 2009b) which is thought to index a re-analysis or evaluation of the stimuli when rule-based expectations are violated.

This ERP evidence for a late positivity in response to metrical violations has been supported by fMRI evidence. Whilst fMRI evidence for the processing of beat or metre in speech is sparse, correctly stressed words have been found to result in areas involved in lexical retrieval (posterior left angular gyrus) and declarative memory (retrosplenial cortex). Stress violations in speech is associated with activity of brain regions involved in phonological (re)analysis (superior temporal gyri, bilaterally), and there are also differences in mild vs. severe metric violations: mild stress violations result in activation of areas involved in higher-order linguistic representation processing (Broca's area and its LH homologue), severe stress violations more recruit activity associated with phonological analyses (left superior temporal gyrus) (Domahs, Klein, Huber & Domahs, 2013). This suggests that a mild stress violation may rely on a top-down input to compensate for the violation and to aid comprehension, whilst severe stress violations result in a more low-level processing difficulty at earlier stages of speech perception.

Taking this evidence together, it seems that when a predictable temporal framework for speech is provided, speech processing is enhanced, whilst violations of expectations results in an engagement of additional resources, indexed by an early negativity followed by a late positivity. To re-iterate a brief consideration regarding native language: temporal cues may be more or less utilised depending on the temporal characteristics of that language. For instance, in a language such as French where stress placement is consistently word/phrase-final, metrical cues may be used less since they do not provide additional lexical information (Cooper et al., 2002).

In conclusion, not only have the temporal aspects of speech been found to interact with speech processing, and at several linguistic levels, but regularised speech may also act to enhance speech processing. In this case, it may be that regularised speech could be a beneficial therapeutic intervention for those suffering from speech and language impairments (e.g. Stahl et al., 2011, 2013). In conducting such studies, however, it is important to consider how findings about speech perception (which have been outlined here) may predict findings about speech production. For this, an understanding of the links between speech perception and production is required.

2.2.3 Speech Perception and Production

The link between speech perception and production is said to be an innately specified relationship (Liberman & Mattingly, 1985). We are said to rely crucially on this interaction during speech acquisition, since auditory feedback of self-made speech allows us to make motor adjustments necessary for target speech production. Indeed, the perception of self-made speech strengthens audio-motor associations required for correct phonemic articulation (Callan, Kent & Guenther, 2000; Guenther, 2006), thus resulting in a parallel, interdependent progression of speech perception and production during normal speech acquisition.

Even in mature listeners, speech production is adapted when a discrepancy between the 'intended' and 'actual' production is detected (assessed by auditory feedback) (Cooper, 1975). This is true for both the production of vowels (Elmen, 1981) and consonants (Shiller et al., 2009). Furthermore, this motor adaption also feeds back onto perception: through a re-learning of audio-motor mappings, a re-categorisation of speech sounds can occur (Shiller et al., 2009). This reveals the cyclic nature of speech perception and production.

The Motor Theory of speech (Liberman, Cooper, Shankweiler & Studdert-Kennedy, 1967; Liberman & Mattingly, 1985) further proposes that speech perception comes about not simply through hearing an auditory signal, but through an activation of motor areas involved in the intended articulation, such as movements of the vocal tract, tongue, lips and jaw. In other words, to perceive speech is to verify the pattern of intended speech gestures. Evidence for these sensorimotor interactions in both speech perception and production has been supported by behavioural, neuroimaging and TMS studies (e.g. Fadiga, Craighero, Buccini & Rizzolati, 2002; Liberman & Whalen, 2000; Watkins, Strafella & Paus, 2003). Subsequent, extended theories also consider motor activation as an integral component of perception, in speech as well as in the broader domains of cognitive neuroscience (see, Galantucci, Fowler & Turvey, 2006).

The link between speech perception and production may go some way to explain the tendency for speech to induce motor entrainment in the form of temporally synchronised speech. These tendencies are most notable in social interaction. Speech perception in conversational contexts may require a temporal synchronisation at several levels: syllabic timing, pauses, breaths, swaying, rate, and in giving feedback to one another in the form of timely interjections (backchannelling) to communicate how they are understanding one other (Bull, 1994; Gill, 2012). Turn-taking has been associated with motor cortex activation, said to reflect a specific role in sensorimotor information processing (Scott, McGettigan & Eisner, 2009, see Figure 2.6).

The motor synchronisation which accompanies speech manifests itself not only as a temporal synchronisation between speakers, but is also evident in the tendency for speakers to produce temporally synchronised gestures alongside speech. Speech gestures are often synchronised to 'beat' of speech being produced (Krahmer & Swerts, 2007; Leonard & Cummins, 2011). These 'beat' gestures can additionally impact on the perception of speech; when gestures coincide with speech prominences they are said to ease comprehension

(Krahmer & Swerts, 2007). A speaker may also begin a gesture prior to speech which can provide a listener with expectations about what, or when speech events will occur (McNeill, 1992; Kendon, 2004). Notably, whilst entrainment (of both gestures and speech) to heard speech (either produced by another speaker or by oneself) does require temporal predictions to be made, it does not require periodicity in the signal. To put it more simply, ‘isochrony’ is not necessary for motor speech entrainment to occur.

2.2.4 Introduction to Chapter 3: Cross-domain comparisons of temporal structures

The next Chapter will consider the relationships between temporal structures in music and speech. As may be clear from the current chapter, temporal structures in music and speech not only share acoustic correlates, but, more importantly, they also seem to share perceptual correlates. As outlined in the introduction, these similarities may indicate there to be common processing mechanisms for temporal structure. This forms the basis of one hypothesis, which is that musical rhythm may impact on speech processing (Figure 1.3). A shared system could have its roots in neural resonance. This may explain the phenomena of attentional orienting, or hierarchical sequencing for example; processes which exist in both domains. These hypotheses will be fully described in the text to come.

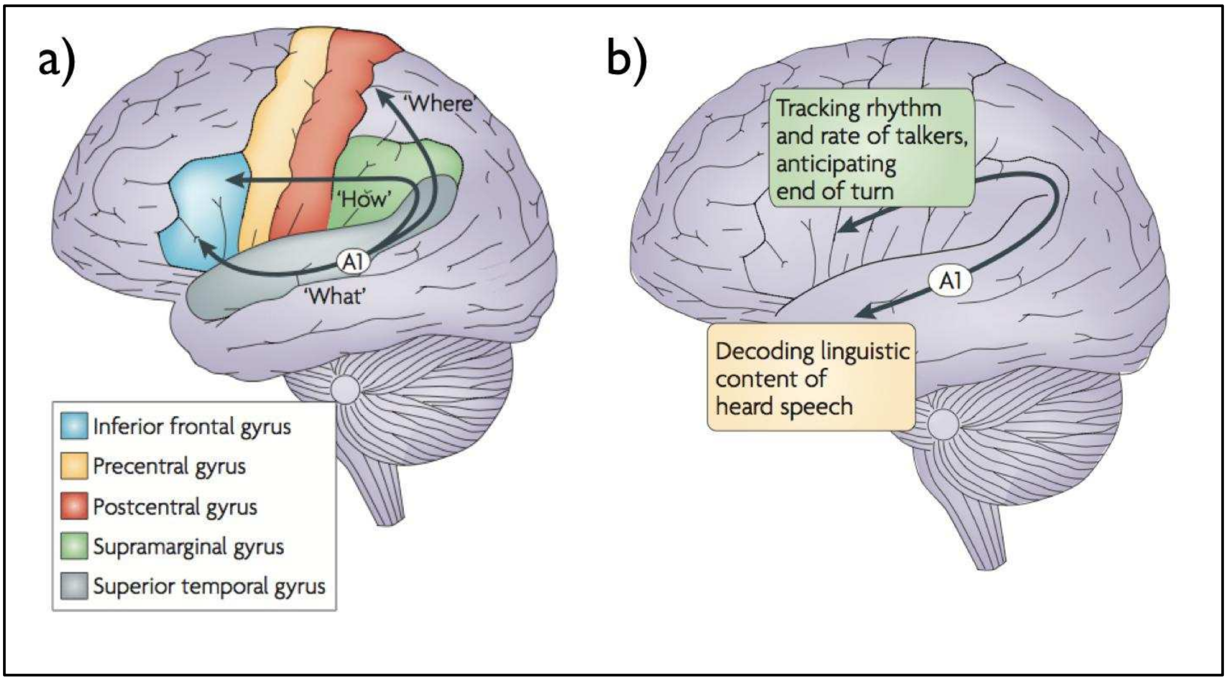


Figure 2.6
Scott et al. (2009)

a) A neuroanatomical model of sound perception as proposed by Scott & Johnsrude (2003): from the primary auditory cortex (A1), anterior streams process meaning ('what') and posterior streams to perform sensorimotor integration ('how' and 'where').

b) Considering its proposed role in sensorimotor integration, the temporal tracking of speech in conversation is proposed to be mediated by the posterior 'how' pathway. This tracking of rhythm and rate is said to enable a listener to anticipate the end of a speaker's turn.

Chapter 3

TEMPORAL STRUCTURES: Cross-domain comparisons

This chapter will attempt to argue that, due to similarities between temporal structures in music and speech, the underlying mechanisms involved in temporal processing may be shared. Indeed, considering that the brain is optimised for efficiency, it seems there would be no advantage in having entirely speech-specific processing mechanisms (Kluender & Greeberg, 1989). Though the functions of speech and music may differ, evidence suggests that neural resources involved in structural integration may indeed be shared and that, by all accounts, this may be true for other cognitive processes (Patel, 1998; Patel, 2012). The idea that cognitive processes involved in language processing are not specific to speech is not a new one (e.g. Langacker, 1987). Here, the idea is that both speech and music draw upon common temporal processes.

The temporal structures described thus far (beat, metre and durational patterning) are considered to involve both automatic and cognitively controlled processes (e.g. Grahn, 2012) which require both referential (e.g. beat intervals) and hierachical (e.g. metrical patterns) relations to be made. In both speech and music, neural resonance has formed an explanation about how we are able to focus attention dynamically in time during online processing. Temporal structures in both domains may also draw upon representational processes. These possibilities will be discussed further below.

The idea that speech and music may draw upon common temporal processes implies that engagement of common temporal processes through music might impact upon speech processing (and theoretically, vice versa). This hypothesis forms the basis of rhythmic priming paradigms, whereby a listener is provided with a temporal context (and thus forms temporal predictions) through musical rhythm. Speech which is subsequently presented - and which conforms to a listeners' predictions - is thought to have a processing advantage compared to speech which does not conform to expectations. Whilst a rhythmic prime could also potentially be provided by speech itself, the greater temporal predictability afforded by musical rhythm (and thus stronger synchronisation tendencies, see Dalla Bella, Białuńska & Sowiński, 2013) may render this priming effect more striking.

In order to speculate about the possibility that temporal processing resources may be shared in music and speech, it is first important to consider the similarities between temporal structures in the two domains.

3.1 Similarities in Temporal Structures in Music and Speech

Both speech and music are hierarchically organised, rule-based systems whose perception rely on the temporal unfolding of acoustic events (Lerdahl & Jackendoff, 1983). Sections 2.2.1 and 2.2.2 have already presented the analogies which can be made between temporal structures in speech and music perception: 'beat' can refer to the perception of isochronous, regular elements, 'metre' can refer to the hierarchical patterning of salient and less salient elements, and durational patterning has been used to refer to the durational variability between constituent components (usually vowel onsets in speech and sound onsets in music).

In this section, the physical and perceptual analogies between speech and music will be described. Following this, the cross-domain effects of music on speech, and motor behaviour associated with temporal structures will be considered. Since temporal structure is considered primarily as a perceptual phenomenon, the next section on physical analogies serves mainly to underline this point, as well as to simply outline empirical comparisons which have been made between the 'physical' structures of speech and music.

3.1.1 Physical Analogies

As already described, the search for physical isochrony in speech has proved somewhat futile (Roach, 1982; Dauer, 1983). Similarly in music, true isochrony of musical performance does not exist; music invariably presents non-isochronous beat intervals in order to accommodate for expressive rubato or tempo changes. Beat perception (isochrony) is therefore considered in both domains to be a perceptual phenomenon (Large, 2008; Lehiste, 1977). Indeed, Ramus, Nespor & Mehler (2000) themselves suggest that isochrony in speech may be a more "abstract construct, possibly similar to that which governs the relation between underlying beats and surface rhythm in music."

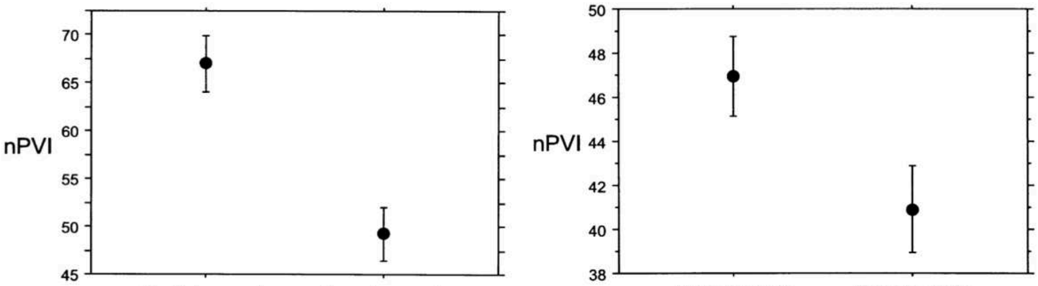
In both domains, however, acoustic signals should have a certain degree of periodicity in order to facilitate beat induction: musical patterns are better perceived and remembered when their tempo (beat intervals) are less variable (Essens & Povel, 1985), for instance. Beat fluctuations are more variable in speech than in music, however, which is supposedly why music is able to induce neural and motor entrainment to a greater degree (Dalla Bella et al., 2013). Acoustic correlates for beat may also be similar in the two domains: the p-centre in speech (to which we assign the 'beat') has been said to display the same acoustic features which induce beat synchronisation in music (Scott, 1998)..

Metre perception in both domains is also largely a perceptual phenomenon: in music this may take the form of subjective rhythmisation, whereby identical isochronous events can be perceived as being more or less salient in order to construct the percept of metre (e.g. Bolton, 1894). In speech, this takes the form of (in an unfamiliar language) hearing speech stresses where there are none (Cutler et al., 1986, 1992; Otake et al., 1993) or not hearing stresses where there should be (Dupoux et al., 1997, 2008; Schmidt-Kassow, Rothermich et al., 2011). This is proposed to be the result of metrical expectancies based on long-term experience within a particular native language. Though the acoustic correlates of stress (an increase in intensity, a lengthened duration or a pitch change) is common to both domains, beat and metre can be primarily considered as perceptual phenomena which can only in part be explained by the raw auditory signal. This leaves the consideration of durational patterning.

The nPVI is a metric which can measure the durational variability of both speech and music. It has thus proved to be a valuable tool in exploring how the durational patterning of music and speech may be related. nPVI values from speech can be calculated by computing the durational difference between one vocalic interval with the previous and measuring it's distance from the mean of this pair (Grabe & Low, 2002; Ramus, 2002). In music, nPVI values have been calculated in much the same way; the interval durations are assigned a score relative to the first duration and these values are inserted into the nPVI equation (Huron & Ollen, 2003; Patel & Daniele, 2003). The nPVI can thus measure the variability between successive intervals in both speech and music. This has provided researchers with an empirical tool to assess the rhythmic relationship between speech and music. Notably, it has been found that the durational variabilities present in language are reflected in the durational variabilities of that culture's music (described next) (Huron & Ollen, 2003; Patel & Daniele, 2003, Figure 3.1). Of course, a measure of the ‘raw’ (speech signal) or ‘intended’ (musical score) auditory signal does not necessarily measure how durational patterning is perceived.

Leaving aside categorisation of languages into rhythmic classes, it has been found that languages do indeed have characteristic durational variabilities. French, for instance, has a lower durational variability (as measured by the nPVI using vocalic intervals) than does English (Grabe & Low, 2002; Ramus, 2002); in other words, the interval durations between adjacent vowels are less variable in French. In two successive studies, Patel & Daniele (2003) and Huron & Ollen (2003) used the nPVI metric to assess whether durational variability in French and English speech was reflected in the respective musics of that culture. Indeed, it was found that French music has a lower durational variability than does English music which well-reflects findings about the durational variabilities of their respective languages. Overall, this evidence was taken to support the observations that native language of a composer impacts upon the durational patterning of their compositions (Abraham, 1974; Wenk, 1987), possibly through an internalisation of the durational patterning of native language which composers unconsciously draw upon.

To conclude, the fact that there are physical similarities between temporal aspects of speech and music may suggest that these aspects are processed similarly in the brain. Whilst the perception of temporal structures is surely reliant on physical sound characteristics, it is additionally a product of temporal expectations, which may derived from an implicit experience-based knowledge (e.g. metrical properties of one’s native language) or domain-general principles of cognition (e.g. hierarchical organisation). Some perceptual analogies between speech and music will now be considered.



Linguistic nPVI values (from Ramus, 2002)
English: mean = 66.99, SE = 2.89, n = 20
French: mean = 49.27, SE = 2.78, n = 20

Musical nPVI values
English: mean = 46.91, SE = 1.81, n = 137
French: mean = 40.90, SE = 1.97, n = 181

Figure 3.1

Patel & Daniele (2003): nPVI values of speech (left, taken from Ramus, 2002) and music (right). The authors found that durational variabilities are higher in English than French musical themes, consistent with what is known about durational variabilities in the respective languages.

3.1.2 Perceptual Analogies

The structured and timely unfolding of auditory events is a key feature of both music and speech, and allows a listener to anticipate both 'when' and 'what' information will occur. Here, the perceptual similarities between temporal structures in speech and music will be considered, bearing in mind the hypothesis that such similarities may indicate a sharing of temporal processing resources.

One similarity between music and speech perception is the detection of an underlying, regular pulse: beat perception. Beat perception has been considered to reflect a domain-general ability to detect irregularities in acoustic signals (Honing, 2011; Honing, 2012), and, in both domains, to induce a dynamic allocation of attention, potentially through a mechanism of synchronised oscillatory activity. In both speech and music, attentional rhythms are said to become attuned to the temporal context and, as a result, there is a preferential processing for auditory events which occur within the attentional window.

In speech, the ABH (Pitt & Samuel, 1990a) claims that attention is directed to points in speech which are expected to be metrically salient (pertaining to the 'beat' in speech). For instance, presenting a listener with a list of words which have an iambic (weak-strong) metrical structure will result in an expectation for this pattern to continue. A similar account exists for the perception of temporal structures in music: The DAT proposes that attention is dynamically allocated to salient musical events (Jones & Boltz, 1989; Large & Jones, 1999). For metre, this model proposes that attentional focus amounts to greatest energy at the most salient events (beat level), as well as a greater attentional energy at strong metrical positions (such as the sub-beat) compared less important metrical positions (see again Figure 2.3, whereby y-axis values could also represent attentional energy). In both domains, when temporal expectations are met, auditory (speech/music) events occurring at the anticipated metrical postition (i.e. within the attentional window) are processed more readily than when they do not (e.g. Bausenhart, Rolk & Ulrich, 2007; Jones, Boltz & Kidd, 1982; Jones et al., 2002). This is thought to be due to the fact that temporal attending can economise the resources available for stimulus processing.

Attentional allocation through music-based beat and metre perception can also cross-modally enhance the detection of visual events (Bolger et al., 2013; Brochard, Tassin & Zagar, 2013; Grahn, 2012b) and can induce a stronger sense of beat in visual beat perception (Grahn, Henry & McAuley, 2011). Though beat can be evoked through visual (Grahn et al., 2011; Grahn, 2012b) and tactile (Brochard, Touzalin, Després & Dufour, 2008) stimuli, the auditory system is said to be specialised for beat perception due to its greater efficiency in encoding temporal precision (Grahn et al., 2011; Patel, Iversen, Chen & Repp, 2005).

Attentional orienting and the subsequent impacts on processing efficiency is thought to come about through an entraining of endogenous neural oscillators to exogenous auditory events: to the musical beat in music (Large & Jones, 1999; Large & Palmer, 2002) and to stressed syllables in speech (Barbosa, 2007; Port et al., 1995). Neural oscillations may also resonate with the less salient auditory events (resulting in oscillations tuned to metrical patterns) in music (Ellis & Jones, 2010; Jones, 2009; Large & Palmer, 2002) and speech (Arantes & Barbosa, 2010; Giraud & Poeppel, 2012).

These complementary oscillations can supposedly interact with, modify, and ultimately strengthen beat expectations. Synchronised gamma band activity (20 - 100 Hz) - associated with attentive focus and spontaneous, synchronised neural firing (Jensen, Kaiser & Lachaux, 2007) - has been said to mark this oscillatory activity in both music (Snyder & Large 2005) and speech (Giraud & Poeppel, 2012). That listeners synchronise to temporal patterns in not only music, but also in speech (Lidji et al., 2011a, 2011b) may complement these findings (in that oscillatory activity may form the basis of motor entrainment).

The perception of metrical patterns is important in performing what can be called ‘rhythmic grouping’ in music and ‘speech segmentation’ in speech. In both cases, auditory events are grouped together (thus also creating perceived boundaries) according to their metrical properties. The fact that auditory segmentation of this kind occurs in both domains also seem to reflect general organisational properties of the brain. Indeed, auditory grouping has been compared to perceptual grouping in the visual domain, whereby items are grouped according to the similarity and proximity, for example (Aksentijevi et al., 2001; Deutsch, 1999; Bregman, 1990, 1994).

As to the neural basis of temporal processing in the two domains, the picture is less clear. Temporal prediction is said to play a central role in speech perception and production, and to involve a subcortical-cortical loop in which is a) specialised for temporal processing and b) which is also involved in the structuring of motor (speech) sequences (Kotz & Schwartze, 2010). This involves interactions between ‘pacemaker’ areas required for providing a basic temporal structure (preSMA, BG and CB) and areas which uses this temporal structure to guide articulation (SMA, premotor cortex, primary motor cortex). Notably, these ‘pacemaker’ areas seem to also be common to music, and are recruited during musical beat and metre perception (see Grahn, 2012).

In all, there may be similar underlying attentional processes which are involved in the processing of temporal structures in both speech and music. However, musical temporal structures are more regular, both in terms of beat intervals and metrical sequences. Indeed, conversational speech metre can be considered to be ‘heterometric’, in that there may be isolated pockets of metrical regularity, but this can change throughout the course of an utterance (from duple to triple metre, for example) (Brown & Weishaar, 2010). In this case, it may be that musical temporal structures, which are inherently more repetitive, recruits more efficiently these attentional mechanisms and, as such, induces stronger temporal expectations. This has formed an explanation as to why music (temporally more regular) is better able to induce motor synchronisation compared to ‘natural’ (temporally less regular) speech (Dalla Bella et al., 2013; Lidji, Palmer, Peretz & Morningstar, 2011a).

The next sections will address evidence for cross-domain impacts of music on speech. Before doing so, it is interesting to consider briefly the relationship between sensory and cognitive processes, particularly in light of EEG research. These considerations are noteworthy when considering at what levels music may impact upon speech.

How cognitive processes come about is a central question in cognitive neuroscience, and one which remains largely unanswered. Whilst sensory encoding involves localised synchronous neural activity, cognitive processing is thought to involve a more global neural synchrony which can be marked in EEG by spectral analysis of theta (6 Hz, linked to memory encoding), gamma (40 Hz, involved in memory, attention and consciousness) and alpha (10 Hz, linked to general cognitive competences and attention) band activity (review Ward, 2003). In ERP studies, cognitive processes may be marked by later latency components, such as the P300, N400 and P600 (Brandeis & Lehmann, 1986). The extent to which sensory encoding contributes to cognitive processes in the auditory domain remains largely unknown. Whilst on the one hand, it may intuitively seem as though higher-order cognitive processes rely on the faithful, initial encoding of sound, there also seems to be an independence (or even top-down effect) of cognitive processes on sensory encoding. For instance, acoustic correlates of metre are not actually required for metre perception.

A link between sensory input and cognitive development has also been made in studies with hearing-impaired individuals (e.g. prelingual deafness): a deprived sensory input in early life means both cognitive and motor sequencing abilities are not able to develop as they should (Conway et al., 2009; Conway & Christiansen, 2009; Horn, Pisoni & Miyamoto, 2006). Nonetheless, the extent to which cognitive processes interact or rely on the sensory encoding of sound remains somewhat of an open question.

3.2 Cross-domain Effects

One way to investigate a possible sharing of temporal processing resources is to study transfer effects from one domain to the other. These may be ‘short-term’ (immediate) cross-domain effects, such as the effect of musical listening on the online processing of speech, or may be a result of long-term experience, such as the effect of musical training on speech functions. Such investigations have contributed to claims that the temporal processing of music and speech share processing mechanisms.

As described in Section 1.2, there are several early levels at which music and speech information encoding is shared. Under the right conditions (see Patel., 2011), musical training can induce a top-down plasticity of subcortical structures, which consequently enhances the phase-locking (frequency encoding) to spectral (pitch) information in speech (Wong et al., 2007) as well as the neural synchrony (timing onset) to speech sounds (Musacchia et al., 2007). It is argued that the sensory encoding of sound is not the only domain-general process which can benefit from musical training, and that cognitive processes (such as auditory attention and working memory) can also be enhanced (Besson, Chobert & Marie, 2011; Patel., 2012; Strait & Kraus, 2011).

One aspect of stimulus processing which may be vital for cognitive operations is temporal processing. One candidate for a domain-general process is attentional allocation through neural resonance. Should this be a domain-general process (and, indeed, modality-general process), one might expect to observe an impact of dynamic attending (via music) on the online processing of speech.

One reliable method for tracking the online response to auditory events is EEG. EEG measures the online responses to sound from the cortex to the order of milliseconds. Of most interest here are the ERPs generated in response to speech, since this allows us to infer the transfer effects from music to speech processing. The level of linguistic processing may be inferred from the latency of the observed component: earlier responses such as the FFR, N19 or P30 are thought to mark auditory brainstem activity (Greenberg et al., 1987) which is involved in sensory encoding of sound. Sensory processing of speech is also marked by early components such as the P50, N100, MMN and P2 which originate from the auditory cortex (Liégeois-Chauvel et al., 1994).

Later components may index the more cognitive stages of speech processing such as stimulus evaluation (P300), semantic-related processes (N400) and syntax-related processing (P600). Musically-induced equivalents have been proposed. This is the case for semantic-related responses (N500, originating from the posterior temporal cortex) (Koelsch, 2005; Koelsch et al., 2004; Steinbeis & Koelsch, 2008). It is also the case for responses thought to mark structural integration in syntax-related processes (P600, Early Right Anterior Negativity, originating from structures involved in the hierarchical sequencing of auditory information, such as the frontolateral cortex, ventrolateral premotor cortex and anterior part of the superior temporal gyrus) (Koelsch et al., 2005; Koelsch, 2005; Maess, Koelsch, Gunter & Friederici, 2001; Patel et al., 1998). The fact that these music-induced responses show similar characteristics to those generated by speech - in terms of both ERP onset latency and their response to violations - may indicate that they represent processes which are general to both domains.

3.2.1. The effect of musical training on speech: temporal structures

The potential for musical expertise to impact on speech may reflect the fact that music engages domain-general processes. Temporal processes engaged by music may also be domain general, in which case one would expect musicians to show an improvement in temporal processing not only in the music domain, but also the speech domain (e.g. detecting speech temporal irregularities), and perhaps even across modalities (e.g. temporal processing of tactile input).

In terms of faithful neural timing to sound onsets of both speech and music, musicians show an advantage over non-musicians (Musacchia et al., 2007). This is also true in terms of hierarchical and sequential temporal processing. Whilst the effect of musical training has already been outlined, here this evidence will be considered in terms of its potential impact on speech processing. This will be considered in cohort with studies directly investigating musical training effects on speech processing.

As described in Section 2.1, musicians are thought to have an enhanced auditory cortex efficiency in processing temporal information in music, as well as a left-lateralised MMN response to metrical deviants (rather than right-lateralised responses as seen in non-musicians) (Vuust et al., 2005): this could imply a musician advantage in speech processing. For instance, professional drummers recruit areas involved in linguistic syntax processing

(LH supramarginal gyrus) when processing complex musical metrical patterns (Herdener et al., 2012).

Electrophysiological research is more abundant. For instance, musician children show a greater MMN response to duration violations of both speech and musical stimuli (Milovanov et al., 2009). Whilst this study cannot measure the relative contribution of musical training (since musician children also had a greater performance on foreign language skills), it demonstrates a link between musical aptitude and speech processing, and at the level of the auditory cortex. The effects of musical training on speech perception has been longitudinally investigated by Chobert et al. (2011), from which the specific effects of musical training can be inferred.

In this study (Chobert et al., 2011), it was found that children who had undergone musical training had a more efficient pre-attentive encoding (than non-musician children) of duration and voice onset time (VOT) deviants of spoken syllables. This was indexed by MMN responses (in a pre-attentive task) as well as behavioural responses (in an active version of the task). This finding has been reproduced by Kühnis et al. (2013) who found that VOT and duration deviants in speech elicit greater MMN responses in adult musicians.

Musicans also better detect metrical incongruities of syllable saliency, as indexed by the P200 component (Marie et al., 2011). In this study, musicians and non-musicians heard sentences which ended on either metrically congruous or incongruous words. A larger amplitude P200 component in response to metrical incongruities in musicians (compared to non-musicians) indicates that musical training results in a greater ability to detect metrical irregularities in speech. Moreover, this study found an interaction of metrical structure with semantic processing of the target words, as indexed by the N400 component (an effect which was, additionally, enhanced in musicians). This may demonstrate the contribution of temporal structure in linguistic functions. The greater ability of musicians to detect speech in noise, which disrupts neural synchrony (Anderson et al., 2010; Parbery-Clark, Skoe, Lam & Kraus, 2009) may also be due to an enhanced sensitivity to, and thus an enhanced ability to attend to the metrical structures of speech (Zion Golumbic et al., 2012).

The ability to detect metrical patterns in speech is important for speech segmentation, a process which is particularly important for language acquisition. In a longitudinal study (François, Chobert, Besson & Schön, 2012), it was found that only children who underwent 2 years of musical training (vs. children who underwent 2 years of painting training) showed an improvement in speech segmentation of an artificial language. This was indexed by a greater amplitude N400 in response to unfamiliar words (Figure 3.2). The positive effect of musical training on speech segmentation has also been found in adults (François & Schön, 2011). In this study, musicians and non-musicians were presented with a sung artificial language, within which statistical regularities allowed them to implicitly extract 'words' from the continuous sung syllables. Similarly to François et al. (2012), it was found that musicians exhibited a larger N400 response to 'non-words' of a sung artificial language, indicating that they recognised the familiar word structure better than non-musicians, and had thus segmented the speech stream more efficiently. This could be due to either an enhanced sensitivity to pitch or an enhanced regularity extraction/sequencing ability (François & Schön, 2011).

Recent evidence also shows that musical training enhances the ability to learn a rhythm-based articifical language (Shook, Marian, Bartolotti & Schroeder, 2013). This study found that musicians were better able to extract the statistical probabilities required for the learning of novel morse-code-based words. This study not only supports evidence of a musician-advantage in learning statistical probabilities (e.g. François & Schön, 2011), but may also demonstrate the sensibility of musicians to the durational properties of sound and their heightened capacity to sequence and segment information.

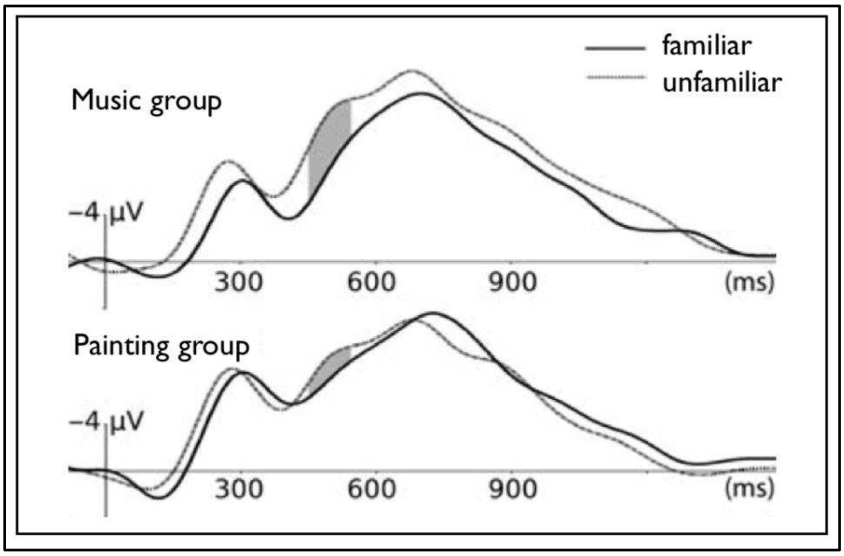


Figure 3.2
François et al. (2012): Musical training enhances speech-segmentation abilities of an artificial language. Figure shows grand averages ERPs to familiar words (black line) and unfamiliar words (grey line). Shaded areas show significant differences (with a larger effect size in musician children).

The prosodic properties of speech (stress, timing and intonation) also provide cues for conveying emotions in speech, the use of which are said to be similar across cultures (Bolinger, 1978). Musicians are not only better able to match a the prosody of a spoken utterance with its musical counterpart (i.e. extract prosodic cues) (Thompson, Schellenberg & Husain, 2003), but are also better able to identify intended emotions in both music and speech (i.e. interpret prosodic cues) (Nilsonne & Sundberg, 1985; Thompson, Schellenberg & Husain, 2004). This is thought to be due to the fact that music relies on similar cues to convey emotional intent (tempo, pitch range, stress contrasts). However, the specific role of temporal cues here cannot be distinguished from pitch cues (prosody being defined by both pitch and temporal fluctuations).

To conclude this section, musical training is thought to fine-tune metrical processing for musical structures (Kung et al., 2011). The evidence presented here suggests that it also enhances the metrical processing of speech (e.g. Marie et al., 2011). The ability to extract metrical cues is required for speech segmentation, and is also enhanced in musicians (e.g. François & Schön, 2011). Musicians also show an advantage in detecting fine temporal differences (VOT) in speech (Chobert et al., 2011). Their greater ability to extract and interpret prosodic cues (e.g. Thompson et al., 2003, 2004) may be related to their enhanced temporal (and pitch) processing abilities.

3.2.2 Effects of temporal orienting on speech processing

The effects of temporal orienting on auditory perception forms the basis of the experiments conducted during this thesis. Whilst there is evidence that, within music and speech domains, priming for beat and metre can enhance auditory processing, the cross-domain impact of musical rhythmic listening on speech processing has not received a great deal of attention. Based on the evidence outlined above it is hypothesised that temporal structures in speech and music engage similar or shared online mechanisms. One candidate mechanism is attentional allocation in time which comes about through temporal orienting: this being the case, it can be predicted that temporal orienting in the music domain (via the presentation of a musical rhythm) engages the same attentional mechanisms to which speech is also sensitive. As a result, subsequent speech could be processed according to the attentional rhythms induced by music. Within-domain effects of temporal orienting will be compared here, followed by cross-domain effects.

Linguistic and musical behaviours may be underpinned by similar temporal processes, namely attention. Since humans are specialised for efficient information processing, it makes sense for us to make use of attention economically and in such a way that we have greater attentional resources available for timepoints at which an event is more likely to happen. The temporal structures of both music and speech are thought to induce synchronised oscillatory activity, and in such a way that temporally expected events coincide with a greater allocation of attention, thus enhancing their processing (Arantes & Barbosa, 2010; Barnes & Jones, 2000; Giraud & Poeppel, 2012; Jones et al., 2002; Large & Jones, 1999; Pitt & Samuel, 1990a).

Within both domains, auditory events occurring ‘on-beat’ are processed more readily than events which occur ‘off-beat’. In music for instance, Jones et al., (2002) found that pitch changes are more quickly detected when they occur at temporally expected position, which implies a more efficient processing. In speech, Quené & Port (2005) found that phonemes are more efficiently detected when they occur at expected beat positions. This effect of beat orienting has been associated with greater cortical and subcortical efficiency in the music domain; Tierney & Kraus (2013a) found that a greater sound processing efficiency at ‘on-beat’ positions was reflected by a greater amplitude P1 response (cortical response) and a greater amplitude brainstem response. In addition, the authors found that the N100 amplitude was greater in response to off-beat events, supporting the idea that it can be elicited in response to a violation of rhythmic expectations. A greater early negativity is also seen in response to metrical violations in speech (for example, if a target word does not adhere to the repetitive metrical patterns of preceding speech) (Rothermich et al., 2010; Magne et al., 2007). Similarities in processing, therefore, are evident from both behavioural and ERP research.

With regards to cross-domain effects (the effect of rhythmic priming on speech processing), we also found that violated metrical expectations (induced by musical rhythm) incurred by speech events results in a greater N100 response (Cason & Schön, 2012 - Experiment 1), and this was accompanied by slower reaction times to phoneme detection. In addition, we also found a larger amplitude P300 component in response to 'off-beat' phoneme targets. A late positivity has also been recorded for temporal incongruities in the speech domain (Astesano et al., 2004; Magne et al., 2007; Schmidt-Kassow & Kotz, 2009a, 2009b) which are, similarly, thought to be elicited when rule-based expectations are violated. We have extended these findings to show that phonological processing of sentences, which present a more metrically complex structure, can also be enhanced by priming with a metrical rhythm (Cason, Astésano & Schön, submitted - Experiment 2). Overall, then, a violation of temporal prediction (cross-domain) mirrors findings from within the music and speech domain, and can be indexed by greater amplitude early negative ERPs and greater amplitude later positive ERPs.

Our findings that temporal orienting through musical rhythm can impact on subsequent speech processing supports some existing evidence reporting a cross-domain impact of musical rhythm to speech. This has been found in particular for the effect of rhythm on lexical access. For instance, a study by Gordon, Magne & Large (2011) found that comprehension of lyrics (lexical decision task: word/non-word) was greater when speech metre was aligned to musical metre. This was also indexed by an increase in induced beta power when there was a metrical alignment, and a delayed peak in evoked gamma response twhen there was a metrical misalignment (strong syllables occurring on weak beats). Induced beta activity and evoked gamma modulations are thought to reflect beat and metre expectancies (Snyder & Large, 2005; Fujioka et al., 2009). As such, the authors interpret the induced beta activity to reflect beat expectations and evoked gamma modulations to reflect the mediation of expectations about metrical strength. In a recent study, Brochard et al. (2013) also found that lexical access is facilitated when visual presentation of correctly segmented syllables (congruent with the syllabic division of the word) occur in line with a simultaneously-presented beat.

The impact of temporal orienting through music on speech processing has also been investigated with clinical populations (Przybylski et al., 2013; Kotz, Gunter & Wonneberger, 2005) which will be discussed further in Chapter 4. The interaction between perception and production in the context of temporal orienting will now be considered.

3.3 Temporal structures and motor behaviour

Whilst temporal orienting may enhance speech perception, how might it also impact on speech production? In addressing this question, it is first interesting to note that there is a relationship between perceptual priming effects and synchronisation abilities. Tierney & Kraus (2013a) found that the effect size of temporal orienting (on cortical responses) is more striking when an individual is better able to accurately synchronise taps to a musical beat. In another recent article, they also show that the ABR is more faithfully phase-locked to an auditory (speech) signal in individuals who have a more accurate tapping ability (Tierney & Kraus, 2013b) (Figure 3.3). This underlines the important relationship which exists between perception and production of temporal structures. Here, the idea that auditory temporal structures in speech and music serve as a template for synchronised motor behaviour will be considered.

Broadly, the term ‘entrainment’ refers to the rhythmic coupling of two independent systems. It has thus far been used to describe neural synchronisation to auditory rhythm, though it can also refer to the motor synchronisation to auditory rhythm. Both neural and motor entrainment seem to require the ability to perceive metrical structure and to coordinate activity with an external stimulus. Though the neural basis of synchronous neural oscillations is not yet clear, it has been implicated in playing a role in motor entrainment, such as synchronisation tapping (Drake, Jones & Baruch, 2000; Tierney & Kraus, 2013a, 2013b): motor entrainment may have its roots in neural entrainment.

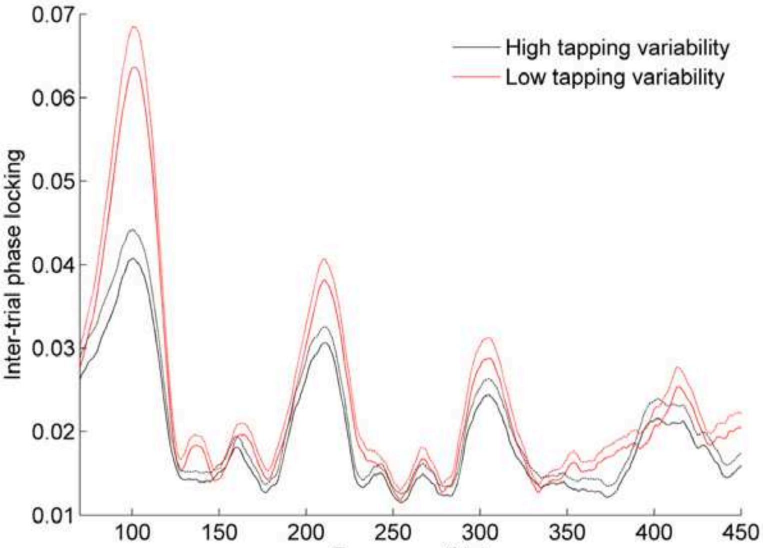


Figure 3.3
Tierney & Kraus (2013b). Phase-locking to the sound /da/ is greater in more accurate tappers.

Auditory rhythmic listening engages motor systems (e.g. Grahn & Brett, 2007) and rhythmic motor production engages auditory areas of the brain (Bangert et al., 2006; Lotze et al., 2003). This co-activation of auditory and motor areas demonstrates the strong audio-motor connectivity in the brain and thus go some way to explain the automacity with which we are able to synchronise our body movements in time to musical rhythm (Drake et al., 2000a). There is also a co-activation phenomenon in speech whereby auditory listening to speech activates motor areas involved in articulatory motor sequences required for its production (Watkins et al., 2003). Importantly, this evidence accentuates the fact that speech (and music) is a multimodal activity which is not experienced uniquely in the auditory modality, and that movement can be considered to be as much a part of speech (and of music) as is sound.

As to the hierarchical levels of synchronization in the two domains, parallels are less clear. In music, motor synchronisation may occur in unison (i.e. 1:1 with the external stimulus), hierarchically (e.g. producing a metrical pattern over a presented beat) or in a polyphonic way (e.g. producing an overlaying rhythmical structure onto an external one) (Clayton, Sager & Will, 2005). Whilst there is little evidence for physical isochrony in speech, speech also has the capacity to entrain movement, either in the form of speech gestures (Krahmer & Swerts, 2007; Leonard & Cummins, 2011) or in the form of synchronised speech (Cummins, 2006, 2009b, in press). Synchronised speech, which occurs in unison, may be guided by not only the auditory prominences or cues of speech (with the p-centre considered as attractors for motor synchronization), but also by the common temporal constraints of our articulatory systems (Cummins & Port, 1998; Port, 2003).

Some studies have directly investigated speech-music similarities in motor entrainment. For instance, both speech and music playing are sensitive to tempo priming (Jungers et al., 2002), and body movement synchronises at similar periodicities in spontaneous speech and improvised music (Gill et al., 2012). Nonetheless, music induces motor synchronisation to a greater degree, and with a greater automacity than does speech. Rather than a music-specific trait, though, it seems to be the greater temporal regularity of music which drives motor synchronisation (Dalla Bella et al., 2013). In both domains, therefore, temporal structures (and, in particular temporal regularity) are able to induce motor synchronisation, be it in the form of sound production (speech/music playing) or body movement, and this may be due to a joint audio-motor representation.

Due to the enhanced audio-motor connectivity of musicians (Baumann et al., 2007, 2005; Zatorre et al., 2007), it might also be predicted that musical training results in an enhanced neural entrainment (see Tierney & Kraus, 2013b). This would presumably enhance the effects of attentional orienting (see Tierney & Kraus, 2013a). In other words, musically trained individuals would be expected to be more sensitive to the cross-domain effect of rhythmic priming as reported in Experiment 1 of this thesis. Indeed, in Experiment 2 (Cason, Astésano & Schön, submitted), we found that even after short periods of rhythmical training, the priming effect (the impact of a metrical prime on phonological processing of a sentence) was augmented. Indeed, even after short periods of training, training can induce neural plasticity (Bangert et al., 2006; Jääskeläinen et al., 2007). However, it remains to be seen whether this augmented effect is due to a greater sensitivity to the musical rhythm (thus a greater attentional orienting effect), or also, additionally, to the speech rhythm itself, independent of the prime.

3.4 Conclusions Chapter 3

This section has described the similarities between temporal structures and their perception in speech and music. Neural resonance to prominent temporal events has been used to account for the attentional effects for both speech and music perception. Whilst the effect of musical training can highlight similarities in processing and implicate specific brain regions or networks, of particular focus is the effects of temporal orienting and attentional allocation on the online processing of sound. This 'priming' effect has been found to enhance the processing of auditory events not only within domains (music-music / speech-speech), but also across domains (music-speech). Priming a listeners' expectations with musical temporal structures may enhance speech processing via an increased temporal predictability of (and heightened attentional resources to) target speech. Whilst temporal orienting may be best induced (and received) in the auditory modailty (particularly through the regularity afforded by music), it may also be both a modality-general and intra-modal phenomenon.

In three studies, we investigated the impact of beat and metre (Experiment 1) or metrical patterns only (Experiment 2 & Experiment3) on phonological speech processing. Interesting to note is that whilst DAT may, to a certain extent, account for findings from these studies, the observed priming effect could also come about through the formation of abstract representations - especially in the case of more complex metrical patterns, through which there is theoretically less potential to induce neural resonance. This will be issue will be discussed in detail in the Discussion section.

Chapter 4

SPEECH AND LANGUAGE PATHOLOGIES: clinical relevance and theoretical support

Here, the potential therapeutic benefits of auditory rhythm for speech- and language-impaired individuals will be considered. In particular, however, pathological studies allow us to test a theoretical idea.

To summarise some of the most pertinent information discussed thus far can be to say that temporal prediction (about both ‘when’ or ‘what’) can enhance speech processing. Temporal predictability may be provided by speech itself, or by musical rhythm. Due to the more regular nature of musical temporal structures, music may induce endogenous oscillatory activity (neural resonance) to a greater degree than speech. This process may act to enhance the perception of speech events. Temporal prediction also forms the basis of motor synchronisation, potentially in form of synchronised production of speech. In acknowledging this potential for temporal structures to enhance both speech perception and production, there emerges an obvious potential for temporal processes to enhance speech in speech- and language-impaired individuals.

Speech impairment can be broadly defined as a compromised ability to produce speech sounds. This term includes articulation disorders (missing or distorted speech sounds), fluency disorders (disrupted ‘flow’ of speech and/or sound repetitions) and voice disorders (abnormal pitch and voice quality). Whilst voice disorders are usually a result of vocal cord abnormalities, articulation and fluency disorders may arise from any number of developmental or neurological causes. Language impairments can be broadly defined as an impairment in understanding and using phonology, morphology, syntax, semantics and/or pragmatics. A language impairment in children can also come about through the poor auditory, phonological and verbal memory skills associated with speech disorders (Sices et al., 2007). Whilst the interactions between, underlying causes and manifestations of speech- and language-impairment is evidently vast, the focus here will be with individuals who suffer from a developmental speech and language impairment, and those who have acquired a speech and language impairment through brain damage.

Developmental speech and language impairments such as dyslexia, developmental stuttering or specific language impairment (SLI) are thought to be caused in part by genetic factors (Bishop, North & Donlan, 1995; Suresh, Ambrose & Roe, 2006; Vargha-Khadem et al., 1998). Speech and language disorders can also result from brain damage or neurodegenerative diseases. Acquired speech and language impairments include non-fluent aphasia, apraxia of speech or neurogenic stuttering, for example.

Notably, some developmental language impairments are thought to arise from a general temporal processing deficit. This has been proposed for dyslexia (Leong & Goswami, 2013; Muneaux et al., 2004; Overy et al., 2003; Thomson & Goswami, 2008), SLI (Corriveau & Goswami, 2009; Weinert, 1992), Pure Word Deafness (Albert & Bear, 1974; Phillips & Farmer, 1990) and individuals suffering from autistm spectrum disorders (who also display language impairment) (Russo et al., 2009). Should this be the case, training with musical rhythm may improve domain-general temporal processing, and, as a consequence, may perhaps ameliorate the associated speech deficit (e.g. Overy, 2000; Kraus & Chandrasekaran, 2010; Tallal, et al., 1996).

Behavioural evidence with dyslexic children may also tell us something about the domain-general nature of temporal processes. In a study by Huss et al. (2011), it was hypothesised that, since metrical processing is important in both music and speech, metrical processing abilities of music should predict those of speech. Specifically, the authors relate this to early, sensory processing. They demonstrate that an impaired musical metre sensitivity in dyslexic children (compared to typically developing children) arises from the same sensory impairment which also effects the recognition of speech metre: an impaired perception of the speech envelope. The low-level temporal coding of speech in this population has also been found to be ameliorated after auditory training, as evidenced by auditory brainstem responses (Russo et al., 2005). It has since been proposed that efficient sensory encoding of regularities and acoustic correlates of temporal structures works in cohort with more cognitive auditory abilities to provide a temporal processing mechanism which is common to both music and language (Strait & Kraus, 2011).

Acquired speech impairments resulting from PD or BG lesions (e.g. dysarthria, non-fluent aphasia) are also associated with a temporal processing deficit (Ludlow, Connor &

Bassich, 1987). This has strongly implicated the BG and SMA as timekeepers not only in speech timing (perception as well as production) (Graeber et al., 2002; Volkmann et al., 1992; Speedie et al., 1993; Stahl et al., 2011; Kotz & Schwartze, 2011) but also in musical time-keeping abilities (Grahn & Brett, 2009; Schwartze et al., 2011). What these findings demonstrate is that a speech-based temporal processing deficit is also associated with a more general temporal processing deficit (including musical ability). Whilst this might tell us something interesting about domain-general processes of speech and music cognition, it is not immediately clear how temporal structures might be used in a clinical context if it is temporal processing which is impaired in the first place. However, it may be input from the CB which can regulate the disturbances in pre-SMA/SMA function (Kotz & Schwartze, 2011). As noted, the CB is engaged by musical temporal structures.

The therapeutic value of 'rhythm' is therefore potentially wide, and this will be considered in the next Section (4.1). Following this, the way in which experimental findings from clinical specch- and language-impaired populations can shed light on temporal processing will be considered in more depth, and in the context of hearing-impaired (HI) children and people who stutter (two populations who have been of interest during the context of this Ph.D).

4.1 Clinical Populations who may benefit from auditory rhythm

This section will outline some of the benefits auditory temporal structures might have in speech pathologies. Movement in persons suffering from motor disorders can benefit from auditory rhythm, notably due to the strong link between auditory rhythm and synchronised movement (McIntosh et al., 1997; Thaut & Abiru, 2010; Thaut, Kenyon, Schauer & McIntosh, 1999; Thaut et al., 1996). Similarly, individuals suffering from speech motor disorders, such as non-fluent aphasia or stuttering may also benefit from auditory rhythmic training or cueing through motor entrainment, possibly via a modified activity of the SMA and BG (both of these pathologies are associated with BG dysfunction). Auditory rhythm also induces neural entrainment, which could be important for temporal and sequential processing - which are important for sound recognition or speech segmentation. Here, the use of 'rhythm' in speech therapies will be outlined for non-fluent aphasic patients and PWS (motor), SLI and dyslexic children (temporal processing) and hearing impaired (HI) children (delayed language acquisition).

Non-fluent aphasia is a speech impairment caused by left-hemisphere damage following stroke. Despite a disruption in the spontaneous expression of speech, many non-fluent aphasic patients are able to produce fluent speech whilst singing. Melodic Intonation therapy (MIT) is a speech therapy which was designed to capitalise on this phenomenon (Albert, Sparks & Helm, 1973; Norton et al., 2009). During therapy sessions, the patient learns to sing syllables at two pitches (stressed syllables at a higher pitch and unstressed syllables at a lower pitch) and to simultaneously tap along with the left hand. This tapping feature of MIT may be a particularly important aspect in some individuals (Dunham & Newhoff, 1979). Therapy sessions progress from the production of single words to short, commonly used sentences, with and then without the therapist. MIT thus consists of a melodic aspect, a rhythmic aspect, and a familiarity aspect, the fluency-enhancing contributions of which have been investigated by Stahl et al. (2011).

This work (Stahl et al., 2011), as well as a more recent longitudinal study (Stahl et al., 2013), presents evidence that singing provides no additional benefit to speech fluency than rhythmic speech alone (also see Boucher et al 2001; Cohen & Ford, 1995). This suggests that the fluency-enhancing effect of singing may be due to the fact that singing is rhythmical. In addition, the authors suggest that the tapping aspect of MIT may engage sensorimotor networks involved in articulatory gestures (see Norton et al., 2009; Gentilucci & Dalla Volta, 2008). Finally, the authors found that patients with larger BG lesions benefited more from rhythmic speech. This research therefore has implications for using rhythm in speech interventions with other populations with BG degeneration, such as PD patients who commonly display monotonic, dysarthric speech (Van Lancker Sidtis, Cameron & Sidtis, 2012).

Using external auditory pacing musical ‘rhythm’ (as opposed to rhythmic speech) is also considered to be an important tool to enhance speech production, through encouraging the entrainment of speech to rhythmic structures. Rhythmic Speech Cueing is described by Thaut (2005), and can include ‘patterned cueing’, whereby each syllable is intoned isochronously with the pacing rhythm provided by the therapist, and/or ‘metric cueing’, whereby strong syllables are produced alongside the strong beat. The success of Rhythmic Speech Cueing is based on the view that auditory rhythm can retrain motor programmes which are also involved in speech production. Rhythmic Speech Cueing is considered to be a potential therapy for non-fluent aphasic, dyphasic, dysarthric and stuttering individuals.

Both patterned and metrical cueing have been found to enhance speech fluency in people who stutter (PWS). Developmental stuttering, primarily a motor timing disorder (Max et al., 2004), is characterised by a involuntary sound repetitions, prolongation of sounds and a difficulty to initiate speech ('blocking') (Ludlow & Loucks, 2004; Yairi & Ambrose, 1999). Notably, PWS produce fluent speech when speaking alongside a metronome (Brady, 1971; Brayton & Contour, 1978; Fransella & Beech, 1965; Toyomura, Fujii & Kuriki, 2011), when regularising prosodic patterns (Packman, Onslow & Menzies, 2000), when speaking in unison, praying (typically rhythmic speech) and when singing (Andrews et al, 1982; Colcord and Adams, 1979; Davidow, Bothe, Andreatta & Ye, 2009; Giraud et al, 2008; Healey et al, 1976; Kell et al., 2009; Webster, 1980). Toyomura et al. (2011) investigated the relative benefits of external auditory pacing in speech fluency of PWS. They found that speaking alongside a metronome (entrainment to 'musical' rhythm) almost abolished stutters, and that speaking in unison (entrainment to others' speech) also somewhat reduced stuttering disfluencies compared to solo speaking. This research forms the basis of a study in which we investigated the role of metrical priming on speech fluency in 8 PWS (Appendix 2).

Speech and language impairments associated with temporal processing deficits are also thought to benefit from training with musical rhythm, supposedly because this training could benefit temporal processing skills which are also important in speech (e.g. Overy, 2000; Russo et al., 2005). SLI and dyslexia are two populations which may suffer from an underlying temporal processing deficit. These impairments are marked by a delayed production of speech and difficulties in speech comprehension, particularly of syntactically more complex speech (O'Hara & Johnston, 1997). Considering that temporal predictability induced by speech metrical patterns can enhance syntactic processing (e.g. Roncaglia-Denissenet al., 2013; Schmidt-Kassow & Kotz, 2009a), temporal predictability induced by musical structures could be hypothesised to do the same. Indeed, the effects of temporally regular auditory stimulation on these two populations has been investigated by Przybylski et al., (2013), who found that rhythmical priming can enhance the syntactic comprehension of spoken sentences. SLI and dyslexic children heard either a 'regular' prime (with an easily-extractable metre) or an 'irregular' prime (with a less easily-extractable metre) and were asked to decide whether a following sentence was syntactically correct or not. Regular primes were found to benefit syntactic processing over irregular primes for both SLI and dyslexic children - as well as for control children. This suggests that short-term listening of temporal structures can focus temporal attention, and that this effect continues even after a delay. Other speech disfluent populations who exhibit a syntactic processing deficit (e.g. patients with BG damage, Kotz, Frisch, von Cramon & Friederici, 2003) have been shown to benefit from the same method of auditory rhythmic listening (Kotz, Gunter & Wonneberger, 2005).

Finally, some methods have been designed to target both the perception and production speech, a relationship which is particularly important in speech acquisition. The Verbotonal Method (VTM), developed in the 1950’s by Guberina, is a method often used with hearing impaired children undergoing speech habilitation [nb. rehabilitation is when an individual is re-learning speech, habilitation is when speech is being acquired for the first time] (Guberina, 1963; Guberina & Asp, 1981; Asp, 1985). HI children experience a simultaneous speech perception and production deficit, and in such a way that an improvement in perception results in an improvement in production and vice versa. The VTM concurs with the view that speech perception and production are interconnected (e.g. Libermann & Mattingly, 1985) and, additionally, that spoken language does not simply involve sound; the whole body is considered to contribute to the perception and production of speech. Children receive simultaneous speech perception and production training (self-produced speech is amplified and/or converted into vibrotactile stimulation which is transmitted to the child through headphones/vibrotactile devices), and they learn to accompany speech with rhythmic movements which intuitively embody prosodic patterning.

Rhythmic training with HI children may also be valuable for the development of sequential and hierarchical processing, in general: early auditory input is thought to be vital for the development of cognitive sequencing abilities (Conway et al., 2009; Conway & Christiansen, 2005). If HI children are unable to receive this from speech, they can instead extract this from tactile cues (Brochard et al., 2008) or more easily-perceived auditory cues (musical rhythm). Furthermore, the temporal cues (>50Hz) alone are enough for accurate speech perception (Shannon et al., 1995): whilst normal hearing listeners also make use also of spectral (pitch) cues, this finding further implicates the therapeutic value of rhythmic cues in speech perception and production in HI children, notably those who use cochlear implants and receive poor spectral (20 - 50 Hz) information.

4.2 Theoretical support for shared temporal processes: HI children and PWS

In particular, studies with pathological populations allow us to test a theoretical idea. In the current context, the theoretical idea is that music and speech processing engages shared temporal processes. An effect from one domain (music) to the other (speech) may indicate an overlap in processing resources.

Specifically of interest here is the cross-domain effect of temporal structures in music to spoken language (speech perception and production). A distinction which has not yet been made is assessing the long-term effects of rhythmic training or cueing (e.g. Stahl et al., 2013) and the short-term effects of rhythmic training or cueing (e.g. Toyomura et al., 2011). Here, it is the short-term, online effects which are of most interest. This is because one candidate for a shared temporal processing does not necessarily require training: online temporal orienting. As noted, however, musical training could enhance the effects of attentional orienting.

During this thesis, empirical investigations into how auditory rhythmic priming can enhance speech production has been investigated in two speech impaired populations described in the previous section: PWS and HI children. The underlying causes and characteristics of speech disfluency in these two populations is very different. However, rhythmic priming can theoretically enhance fluency regardless of this, due to the fact it engages both cognitive attentional (perception) and motor synchronisation (production) processes which may be domain general. For instance, it was hypothesised that for PWS (primarily a motor disorder), musical rhythmic priming would provide a template for speech production and induce speech motor synchronisation of sorts. In the case of HI children (primarily a perceptual disorder), it was hypothesised that temporal predictability afforded by musical rhythm can direct attentional resources in time, thus optimising processing efficiency for temporally expected speech events. Through this enhanced perception (of a target sentence), they would display a better production (of the target sentence).

Whilst studying both populations can inform us more about whether musical rhythm is indeed useful in spoken language performance, we can additionally speculate as to how this might occur, given what we know about the underlying causes of the pathology. For instance, the underlying neural bases of stuttering may tell us more about the role of the basal ganglia in music and speech, and the response to rhythmic priming in HI children may tell us more about how temporal structures are important in language acquisition, or the acquisition of sequencing abilities in general (Conway et al., 2009).

The BG are structures which are recruited during general timing processes (Buhusi & Meck, 2005), and also, more specifically, to beat perception in music (Grahn & Brett, 2007). These may also be important structures in detecting temporal regularity in speech as well as in the planning of timely speech production (Kotz, Schwartze & Schmidt-Kassow 2009; Kotz & Schwartze, 2011). The BG have also been proposed to play a general role in attentional mechanisms which are able to bind input and output information (Brown & Marsden, 1998), perhaps explaining why BG dysfunction is associated with syntactic processing defecit (Kotz, Frisch, von Cramon & Friederici, 2003). Many of these findings come from studies with clinical populations, in whom it is observed that BG dysfunction (through stroke or degeneration) show temporal processing (beat) deficits (Grahn & Brett, 2009; Harrington, Haaland & Hermanowicz, 1998; Schwartze et al., 2011) and temporal production deficits (Harrington et al., 1998). These patients are nonetheless highly responsive to rhythmic stimulation, possibly due to CB inputs which regulate abnormal activity (Kotz & Schwartze, 2011).

Stuttering is also associated with BG dysfunction (Alm, 2004) and altered BG-thalamo-cortical circuitry (Lu et al., 2010), with stuttering severity correlated with BG activity (Giraud et al., 2008). PWS also show a regulation of abnormal BG activity (and a parallel enhancement of speech fluency) during entrainment of speech to an external beat (Toyomura et al., 2011). As noted, music may have the potential to regulate dysfunctional BG circuits. Stuttering has also been associated with deficits in the planum polare and left anterior insula - areas which are important for the processing of metre (Kell et al., 2009; Liégeois-Chauvel et al., 1998; Vuust et al., 2006).

The perception of metrical structures in speech is thought to be a vital prerequisite for speech acquisition, allowing for speech segmentation to occur, as well as perhaps other cognitive processes which require sequential processing. Speech disfluency through HI arises from inability to accurately detect sound. Considering the nature of the auditory pathway (namely, it exhibits a bottom-up and top-down plasticity (Schofield, 2010)), a deficit in one, low level, structure (i.e. the cochlear) will consequently impact upon the development (and strength) of the entire pathway. Indeed, in the absence of bottom-up (afferent) input to the auditory cortex, top-down (efferent) inputs are also unable to mature - as a consequence, this cuts the potential for a top-down plasticity to occur and thus also prevents learning (Kral & Eggermont, 2007; Kral et al., 2000).

Music-induced plasticity of subcortical structures which are also shared for the encoding of speech can occur via these efferent tracts, however (Wong et al., 2007). Considering this, music may be able to, to some extent, provide HI children with inputs they are not able to receive via speech: it has been suggested that early sound input provides a scaffold for general cognitive sequencing abilities (Conway et al., 2009). This explains why HI children thus perform worse than normal hearing children on sequence learning, across modalities (Conway et al., 2009; Conway & Christiansen, 2005). It may be the case that musical metre are able to provide such a 'scaffold', since it would be more easily heard by HI children than would speech. Indeed, HI children can acquire spoken language to the same level of normal-hearing children if intervention is started early, which demonstrates the plasticity of the young brain and thus also the potential for responding to music-induced plasticity.

4.3 Conclusions: Chapter 4

To provide a brief summary: performing research with speech- and language-impaired populations can not only tell us more about the pathology of the impairment itself, and therefore the potential therapeutic benefits of auditory temporal structures, but can give us an insight into how temporal structures are processed in the brain.

The research we have conducting with subjects without a speech / language defecit may tell us about the processing advantages of temporal orienting (Experiment 1, Experiment 2). In order to test whether temporal orienting can also enhance production, it may be necessary to perform studies with individuals who do not have ceiling-performance production skills. In addition to this, given what we know about the neural bases of a speech disorder, we can also speculate as to what candidate areas might be involved in a cross-domain temporal orienting effect.

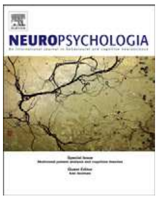
EXPERIMENT 1

Neuropsychologia 50 (2012) 2652–2658

Contents lists available at SciVerse ScienceDirect

Neuropsychologia

journal homepage: www.elsevier.com/locate/neuropsychologia



Rhythmic priming enhances the phonological processing of speech

Nia Cason*, Daniele Schön

U 1106, Institut de neurosciences des systèmes, AMU, Centre MEG, Service de Neurophysiologie Clinique, CHU Timone, Rue Saint Pierre 13385, Marseille, France

A R T I C L E   I N F O

Article history:
Received 4 April 2012
Received in revised form
4 June 2012
Accepted 10 July 2012
Available online 22 July 2012

Keywords:
EEG
Speech and Music
Rhythm
Temporal Expectation

A B S T R A C T

While natural speech does not possess the same degree of temporal regularity found in music, there is recent evidence to suggest that temporal regularity enhances speech processing. The aim of this experiment was to examine whether speech processing would be enhanced by the prior presentation of a rhythmical prime. We recorded electrophysiological (EEG) and behavioural (reaction time) data while participants listened to nonsense words preceded by a simple rhythm. Results showed that speech processing was enhanced by the temporal expectations generated by the prime. Interestingly, beat and metrical structure of the prime had an effect on different ERP components elicited by the following word (N100, P300). These results indicate that using a musical-like rhythmic prime matched to the prosodic features of speech enhances phonological processing of spoken words and thus reveal a cross-domain effect of musical rhythm on the processing of speech rhythm.

1. Introduction

Language and music are two universal human activities. Though both vary hugely across cultures, they are thought to be biological predispositions for which our cognitive and perceptual systems are primed (Trehub, 2001). Both are hierarchically organised, rule-based systems which are dependent on how acoustic events unfold over time (Lerdahl & Jackendoff, 1983). Valuable comparisons between the two domains have allowed us to consider to what extent their processing draws on similar or different mechanisms (Patel, 2008).

Of the several parallels which can be drawn between language and music, of present interest is 'rhythm'. Broadly speaking, rhythm is the temporal organisation of acoustic events which unfold over time, and is important for the perception and production of both speech and music. More specifically, we are interested in metrical structure: the regular alternation of elements that are perceived as 'strong' or 'weak'. In music, the periodicity of strong elements forms the basis for beat perception, whilst the hierarchical organisation of beats allows for the emergence of metre. As a simple example, regular repetitions of 'strong–weak' units form a march-like, binary structure (1 2 1 2 1 2), and repetitions of a 'strong–weak–weak' unit forms a waltz-like, ternary structure (1 2 3 1 2 3 1 2 3). Similarly, this patterning of strong and weak elements occurs in speech, as stressed and unstressed syllables ('metric feet') and offer important prosodic information. In this paper, we work around the supposition that speech perception and production across languages is, like music, dependent on beat (stressed, 'strong' syllables) and metre (the alternation of stressed and unstressed – 'strong' and 'weak' – syllables). Indeed, even in syllable-timed languages such as French, characterised by equal syllable duration, stressed syllables play an important role in the rhythmic organisation of speech (Hirst & Di Cristo, 1998), occurring consistently at the end of a word. As such, it may be argued that binary (1 2 1 2 1 2) and ternary (1 2 3 1 2 3 1 2 3) metrical structures in music may be analogous to a bisyllabic ('weak–strong') and a trisyllabic ('weak–weak–strong') word in speech.

One might argue that making an analogy between rhythm in music and rhythm in speech is problematic, especially since natural speech does not possess the same degree of temporal regularity found in music. On the other hand, listeners may perceive stressed speech events to occur isochronously (Lehiste, 1977; Schmidt-Kassow, & Kotz, 2009), and prosodic information is often repetitive which allows it to be predictable (Dilley & McAuley, 2008; Martin, 1972; Meltzer, Martin, Mills, Imhoff, & Zohar, 1976; Shields, McHugh, & Martin, 1974). This suggests there is a preference for regularity in speech, and that this regularity enhances speech processing (Quené & Port, 2005). This is also evident for speech production, where stressed syllables have a tendency to be produced at regular intervals on repeating spoken phrases (Cummins & Port, 1998). With this in mind, metre perception in speech and music may be dependent on similar mechanisms, and, to take this further, carry-over effects after a regular stimulus has ended (such as entrainment or the continuation of a memory trace) may infringe upon the perception of a following auditory event in a cross-domain manner.

---

* Corresponding author. Tel.: +33 4 91 38 55 77.
E-mail address: nia.cason@etumel.univmed.fr (N. Cason).

0028-3932/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.neuropsychologia.2012.07.018

In order to form temporal predictions about a sequence of rhythmic events, dynamic attending theory proposes that internal neuronal oscillations entrain, or synchronise to, external auditory events (Jones & Boltz, 1989; Large & Jones, 1999; Large & Snyder, 2009). More precisely, internal oscillations are said to 'hook' onto different metrical levels of a rhythm with phase and period relations which correspond to the hierarchical structure of music (Large & Palmer, 2002). Through this mechanism, points at which an event is expected to occur (e.g., at the onset of a strong element) act as a magnet for attentional resources, and as a consequence perceptual awareness at these points is heightened. Similar models for rhythm-directed attention also exist for speech perception, for example, the "Attentional Bounce" hypothesis (Pitt & Samuel, 1990) which claims that attentional resources are preferentially allocated to locations at which stressed syllables are predicted to occur.

A number of studies have investigated this perceptual heightening at attended time-points and its impact on the processing of non-speech stimuli, showing that the processing is better at attended time-points than unattended ones (Jones, Boltz, & Kidd, 1982; Jones, Moynihan, MacKenzie, & Puente, 2002). What follows is the possibility that dynamic attending may also enhance the processing of speech sounds. Investigations into the effect of temporal expectation in speech began by a study conducted by Meltzer et al. (1976). They found that when a target phoneme embedded within a spoken sentence occurred 'on-time' (i.e., no alterations to the speech signal), it was detected more quickly than a target phoneme which was shifted to occur 100 m s either before or after the natural speech timing ('off-time'). In the study by Pitt and Samuel (1990), experiment 2 of 3, rhythmical expectations were induced by presenting a list of words with identical stress patterns. When a target phoneme occurred in a syllable which was predicted to be stressed it was detected faster, presumably because attention was directed toward the salient syllables. In a related study, Quené and Port (2005) found that a target phoneme was more quickly detected when the preceding lists of words were aligned in such a way that their stressed syllable occurred at regular temporal intervals. The authors interpret this as evidence that regular timing of speech enhances speech perception. In all, these results support the view that a listeners expectancy of the timing of stressed syllables contributes to speech perception (Lehiste, 1980; Port, 2003) and that by inducing expectations about speech rhythm, phonological processing can be heightened. On the basis of this, as well as on the similarities of rhythm in music and speech, it seems plausible that a musical-like prime with the ability to induce temporal expectations may also be used to enhance speech processing.

We hypothesised that a musical-like rhythmic prime sequence would induce expectations about both beat position (at regular 700 m s intervals) and about metrical structure (either binary or ternary). The presentation of a target word (following the prime) either conformed to the temporal structure and thus to listeners' temporal expectations, or did not. More specifically, a target phoneme either occurred at the point at which a 'strong' element was expected to occur (henceforth 'on-beat'), or it occurred after this point by aligning the previous syllable, instead, with the point of expectation (henceforth 'off-beat'). Second, the number of syllables in the target word was either congruent or incongruent with the prime's metrical structure (binary or ternary). We expected to find that when beat/metrical expectations were met, the processing of a target word would be facilitated, as indicated by faster reaction times to a phoneme detection task. We measured event-related brain potentials (ERPs) in order to disentangle whether the predicted faster reaction times might be due to sensory (early ERPs) or motor decisional (late ERPs) processes. To this aim, we asked participants to complete a phoneme detection task in 240 trials consisting of either binary or ternary metrical structures followed by a rhythmically matching or mismatching pseudoword.

2. Methods

2.1. Participants

20 participants (six women) between the ages of 21 and 40 years (mean = 27 years, 10 months) participated in the experiment. All participants had normal or corrected-to-normal vision, and French was their first language. Behavioural and electrophysiological data were collected during a reaction-time task (phonemic detection task). EEG data from two participants were excluded due to technical problems and bad signal data (making 20 behavioural and 18 EEG participants). Each participant gave informed consent prior to experimentation and was compensated for their time with 15 euros.

2.2. Stimuli

Experimental trials consisted of a prime sequence (either a Binary prime or a Ternary prime) followed by a pseudoword (either bisyllabic or trisyllabic).

Prime sequences were created in Adobe Audition. Binary primes consisted of a weak-strong structure (350 m s stimulus onset asynchrony, SOA) repeated four times. Ternary primes consisted of a weak-weak-strong structure (233.3 m s SOA) repeated four times. Strong tones had a frequency of 900 Hz, while weak tones had a frequency of 450 Hz and an intensity reduced by 9 dB. Thus the beat per minute was 85.7 (700 m s) in both primes. All tones had a duration of 100 m s with a rise/fall time of 5 m s. These binary and ternary units were repeated four times, resulting in a Binary prime sequence with a duration of 2550 m s, and a Ternary prime sequence with a duration of 2666 m s. Bisyllabic (1 8 0) and trisyllabic (1 8 0) pseudowords were constructed using a pseudo-random concatenation of consonants and vowels (New, Pallier, Ferrand & Matos, 2001). They did not resemble existing French words, though were phonetically and orthographically valid. Bisyllabic words had a consonant–vowel/consonant–vowel ([cv/cv]) structure and trisyllabic words had a [(c)cv/cv/cv] structure). Their last syllable contained a plosive consonant (/p//b//t//d//g/ or/k/), which matched a visually presented target phoneme in 50% of experimental trials. Consonant pairs with similar place of articulation (/d/ and /t/, /b/ and /p/, and /g/ and /k/) did not occur in the same word and were not used in conditions when the target phoneme was not present so that they were not confused with one another. Pseudowords were recorded in a soundproof booth by a native French speaker. For each pseudoword, the duration of syllables was calculated by determining the syllable onsets and offsets (PRAAT speech segmentation software, Boersma & Weenink, 2012). These duration values were used to determine the word onset following the prime, so that the final syllable (containing the target phoneme) would fall either on-beat or off-beat in relation to the beat inferred by the prime.

2.3. Procedure

Participants sat in front of a computer screen. The volume for auditory stimuli was adjusted to a comfortable level. Participants were asked to decide whether the visually-presented target phoneme which would be presented on the screen in front of them was or was not present in the auditorily-presented pseudoword, and to make their responses as quickly as possible by pressing either a 'yes' button or a 'no' button. The hand in which each button was held was balanced across subjects in order to control for handedness.

After a training session of 24 trials (twelve trials for each of the two prime sequences), the experiment began, consisting of two blocks (binary and ternary prime blocks, counterbalanced) of 120 trials each. For each trial, the target phoneme (/p//b//t//d//g/or/k/) was presented on a computer screen (in lowercase) for the duration of the trial, and their presentation was blocked across trials, 20 trials for each of the 6 target phonemes. The order of presentation of target phonemes was counterbalanced across participants. In each trial the prime was followed by a pseudoword. Pseudowords could either match or mismatch the metrical structure of the prime (whilst a bisyllabic word is considered to match a binary prime, a trisyllabic word is considered to mismatch it). Moreover, the target phoneme within the pseudoword could either occur on or off the strong beat generated by the prime (see Fig. 1). The order of experimental conditions (Beat conditions ('on-beat' or 'off-beat') and Metrical conditions ('matched' or 'mismatched') was pseudo-randomised within each experimental block. Participants heard each pseudoword only once. However, words were used in all four conditions across participants (fully balanced). Error feedback was received at the end of each experimental block. Participants were asked to avoid moving their head and to fix their eyes on a fixation point on the computer screen. Both visual target phoneme and auditory stimuli were presented using Presentation Software. Audio examples of the trials are available as Supplementary material.

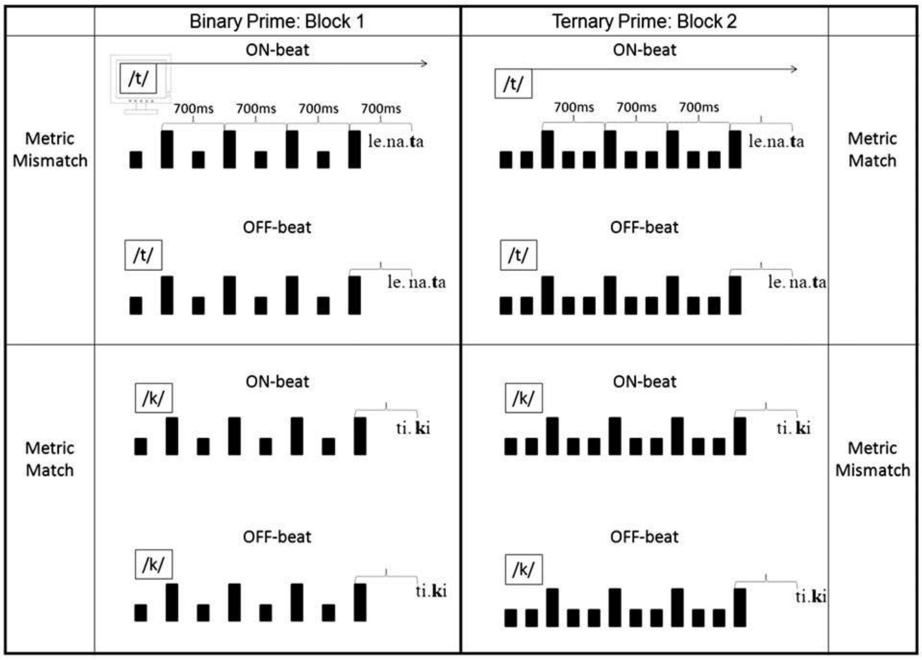


Fig. 1. Counterbalanced Experimental Design: binary primes (left column) and ternary primes (right column) were followed by the presentation of a spoken pseudoword. Its target phoneme (in bold font) was either aligned with beat periodicity of 700 m s (on-beat) or was not (off-beat). In addition the pseudoword was either metrically matching or metrically mismatching the metric structure of the prime (a binary/ternary prime was matched or mismatched by bisyllabic/trisyllabic pseudowords). The target phoneme was always contained in the last syllable of the pseudoword, which is stressed in French. Audio examples of the trials are available as Supplementary materials.

2.4. Data acquisition and analysis

Participants were comfortably seated in a Faraday booth. The EEG was recorded from 32 scalp electrodes (Biosemi ActiveTwo system, Amsterdam University) located at standard left and right hemisphere positions over frontal, central, parietal, occipital, and temporal areas (International 10/20 system sites: Fz, Cz, Pz, Oz, Fp1, Fp2, AF3, AF4, F3, F4, C3, C4, P3, P4, P7, P8, Po3, Po4, O1, O2, F7, F8, T7, T8, Fc5, Fc1, Fc2, Fc6, Cp5, Cp1, Cp2, and Cp6). The bandpass filter was set between DC and 102.4 Hz and sampling rate was 512 Hz. The data were then re-referenced offline to the algebraic average of the left and right mastoids and high-pass filtered (1 Hz). Raw data containing movement artefacts or amplifier saturation were excluded. Signal containing ocular artefacts was corrected using Independent Component Analysis decomposition by which the component containing the blink was removed (Makeig, Belll, Jung, & Sejnowski, 1996). The EEG was then epoched time-locked to the onset of the target phoneme and the 100 m s baseline (before the phoneme onset) zero-mean normalized using Brain Vision Analyzer software (Brain Products, Munich, Germany).

Trials were averaged for each condition and participant and for each condition across participants (grand average). After visual inspection of the grand averages and a time-line analysis with 50 m s time windows, three time windows were chosen for statistical analysis: 0–100, 100–250 and 250–500 m s. While the 0–100 m s latency window acts more as a control window, the 100–250 m s was used to describe activity during an early negative deflection while the 250–500 m s latency window captures the later positive component, the P300.

Analyses were performed for correct trials only. Repeated-measures analysis of variance (RM-ANOVA) was used for statistical assessment of both behavioural and ERP data, including 'Prime Type' (binary or ternary prime), 'Metre' ('matched' or 'mismatched'), and 'Beat' ('on-beat' or 'off-beat') as factors. The analysis of behavioural data additionally included the 'Expected response' (yes/no) factor while the ERP data additionally included 'Laterality' (left/medial/right), 'Antero-Posterior Gradient' (frontal/central/parieto-occipital), and 'Electrode' (three electrodes per region which were separated into these categories by dividing a figure) to model the topographical distribution of the effects (frontal left (F7, F3, AF3), frontal medial (Fz, FC1, FC2), frontal right (AF4, F8, F4), central left (FC5, C3, CP5), centro-medial (Cz, CP1, CP2), central right (FC6, C4, CP6), parieto-occipital left (P7, PO3, P3), medial parieto-occipital (Pz, O1, O2), right parieto-occipital (P8, PO4, P4)). All p-values reported below were adjusted using the Greenhouse–Geisser correction for nonsphericity, when appropriate, and Tukey tests were used in post-hoc comparisons.

3. Results

3.1. Behavioural data

All participants were able to perform the task easily (mean percentage of incorrect/missed trials in the experimental blocks=6.8). Correct rejection responses (correct 'No' responses) were significantly slower than hit responses (correct 'Yes' responses) (mean hit=580.6 ms, mean correct rejection=648.6 m s $F(1,19)=24.65$, $p<.0001$). In addition, there was a 'Response' by 'Beat' interaction ($F(1,19)=16.438$, $p<.001$); post hoc comparisons showed that reaction times to targets presented 'On' the beat were faster compared to targets presented 'Off' the beat (mean=572 m s and 588 m s, $p<.001$) while this was not the case when the target was not present (correct rejection; on-beat mean=650 m s, off-beat mean=647 m s). These correct-rejection trials require a longer searching period that might hinder the detection of subtle on-beat versus off-beat timing differences, and so for this reason, we ran a new analysis (again, on correct responses) of only trials containing the target phoneme (i.e., hits).

A three-way RM ANOVA was run on this reduced behavioural dataset (hits) with the factors 'Prime' (binary/ternary), 'Metre' (matched/mismatched) and 'Beat' (on-beat/off-beat).

The on-beat condition yielded significantly faster reaction times than the off-beat condition ($F(1,19)=13.13$, $p=.0018$). There was no 'Beat' by 'Metre' interaction ($F(1,19)=.364$, $p=.5534$).

There was no significant difference between 'Metre' conditions. However, there was a significant interaction between 'Prime' (binary/ternary) and 'Metre' (matched/mismatched) ($F(1,19)=14.66$, $p=.001$). Post-hoc comparisons revealed that though the difference between matched and mismatched metre conditions was significant within the binary prime block (566 and 531 m s respectively, $p=.042$), it did not reach significance in the ternary prime block (554 and 584 m s respectively, $p=.08$).

3.2. Electrophysiological data

In the 0–100 m s latency window there was no significant main effect or interaction of the experimental factors.

In the 250–500 m s latency window off-beat conditions yielded a significantly more positive ERP component than did on-beat conditions (off-beat.2606 μV, on-beat −.39404 μV, significant main effect of 'Beat', $F(1,17)=17.778$, $p=.00058$). There was also an interaction of this effect with 'Laterality' and 'Region' ($F(4,68)=6.529$, $p=.00016$), due to a greater effect over central sites compared to frontal and parieto-occipital sites, especially over left and midline regions (Fig. 2b).

A second analysis was run on the P300 latency at electrodes P3, Pz and P4, also in the 250–500 m s latency window, wherein amplitude was maximal. Data from two participants in the Binary block and four participants in the Ternary block was removed due to absence of a clear P300 component. There was a trend for off-beat conditions to yield a longer P300 latency (mean on-beat latency=431.276 m s, mean off-beat latency=467.556 m s, $F(1,17)=3.732$, $p=.07$).

In the 100–250 m s latency window, mismatched metre conditions yielded a significantly more negative ERP component than did matched metre conditions (matched value, −1.00817 μV, mismatched value, −1.39821 μV, significant main effect of 'Metre', $F(1,17)=7.713$, $p=.012$). There was an interaction of the factor 'Metre' with 'Laterality' and 'Region' ($F(4,68)=4.328$, $p=.00354$). Post hoc comparisons showed that this arose from the signal difference between matched and mismatched metre conditions at central electrodes ($p=.005$) (whereas there was no significant difference between ERPs yielded by these conditions at frontal or parietal electrodes), and that within this central region, the effect was spread over right, medial and left electrodes with the largest effect seen at the central medial and right electrodes (Fig. 2a).

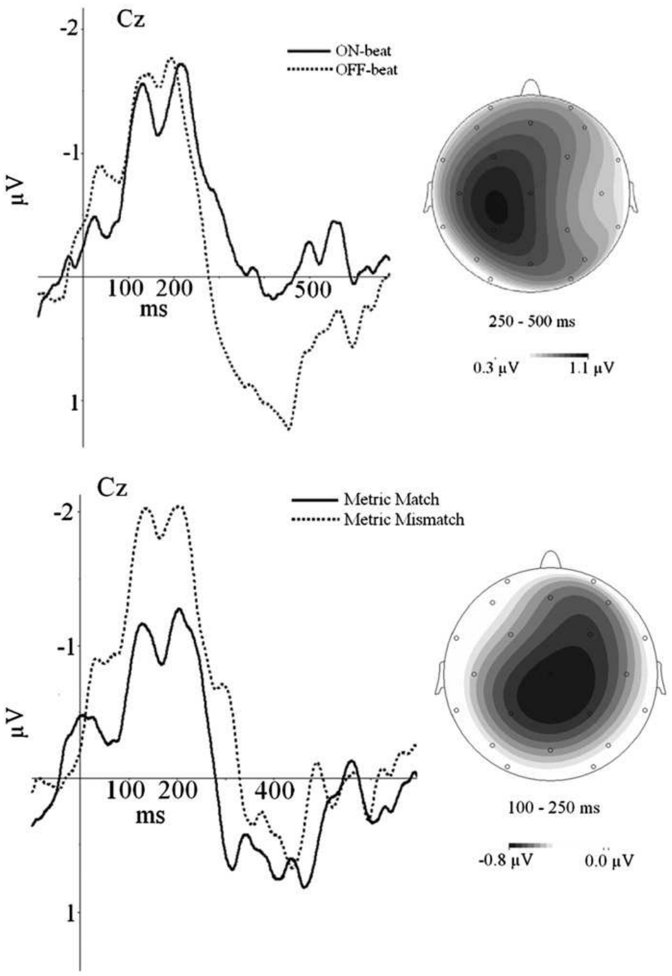


Fig. 2. Top: Grand-Averages Event-Related Potentials to on-beat (dotted) and off-beat (thick) target phonemes (Cz electrode, $N=18$). The map shows the distribution of the effect (off-beat–on-beat) in the 250–500 m s range window. Bottom: Grand-Averages Event-Related Potentials to metric match (dotted) and metric mismatch (thick) target phoneme (Cz electrode, $N=18$). The map shows the distribution of the effect (mismatch–match) in the 100–250 m s range window.

4. Discussion

A musical-like rhythmic prime was designed to induce expectations about both beat and metre. We measured behavioural responses and ERPs to spoken pseudowords which succeeded the prime in order to investigate whether phonological processing is enhanced under conditions in which rhythmical expectations are met. We will first discuss the effects of beat expectancy on phonological processing, and then the effects of metrical expectancy.

4.1. *The effect of beat expectancy on speech processing*

Here, we will discuss the finding that phonological processing was more rapid when beat expectations were met, and will then focus on the ERP component identified, the P300, which predicts these behavioural results.

Phonological processing constitutes the building blocks of speech perception. A target phoneme occurred either at an expected (on-beat) or unexpected (off-beat) position in relation to the prime beat. The fact that target phonemes were detected more rapidly at attended locations supports findings that attentional orienting can facilitate auditory processing. This has previously been shown for events such as pitch changes and melodic violations (Barnes & Jones, 2000; Jones et al., 1982, 2002) as well as for speech sounds (Pitt & Samuel, 1990; Quené & Port, 2005) Interestingly, in the present study, rather than a simultaneous presentation with the 'strong' point, the target phoneme occurred as a continuation of (i.e., after) the prime. Thus the effect was both 'long-distance' (see Dilley & McAuley, 2008) and also cross-domain, insofar as a musical-like rhythmic prime affected the processing of spoken words.

Because auditory rhythmic patterns are known to have a strong effect on motor entrainment (Thaut, Kenyon, Schauer, & McIntosh, 1999), one might ask whether faster reaction times from on-beat conditions could have been due to a tendency to give the response at a closer temporal vicinity to the inferred beat. Two scenarios are possible here. In the first, participants may anticipate the response to make it coincide with the beat where the target phoneme is presented. This synchronisation behaviour would necessarily result in a higher false alarm and error rate in on-beat conditions while this was not the case. Alternatively, one may expect participants to wait for beat following the target phoneme presentation and synchronise to that. In this case behavioural responses would be close to the next metrically strong position following the word, which they were not.

Orienting to regular points in time was also employed by Quené and Port (2005) in a study which investigated the influence of rhythmic structure of words and timing regularity on speech perception. Whereas they induced temporal expectancies through the presentation of bisyllabic words, we used a musical-like rhythmic prime; whilst they found that a regular speech rhythm improves speech perception, we found that using a regular musical-like rhythmic prime is able to do the same. Our study differs from Quené and Port's (2005) in that we measured ERPs of phoneme detection occurring at these temporally expected versus unexpected locations: the differences in the P300 amplitude and latency further show the effect of temporal expectations on phonological processing. Specifically, a beat expectancy violation (a phoneme target occurring asynchronously with the inferred beat) resulted in a larger-amplitude and a tendency for a longer latency P300 response.[1]

The finding that beat expectancy violations evoked a greater amplitude P300 is in line with the classical interpretations of an increased P300 amplitude as being inversely proportional to the expectancy of a stimulus (Duncan-Johnson & Donchin, 1977; Squires, Wickens, Squires, & Donchin, 1976; Tueting, Sutton, & Zubin, 1970), as well as with the context-updating hypothesis (Donchin, Karis, Bashore, Coles, & Gratton, 1986), whereby there is an increase of the P300 amplitude in response to low-probability events within a sequence. Stimulus expectation in our own study was induced by the rhythmic prime, via entrainment, through which an implicit temporal expectation about target location was formed. In this respect, there is, similarly to the classic oddball, a habituation to the regular sequence (the prime), and in response to an unexpected stimulus event (an off-beat target phoneme) a larger P300 amplitude is elicited. Explicit temporal orienting of attention via the presentation of an informative cue is also reported to have a similar modulation of the P300 (Correa, Lupiáñez, Madrid, & Tudela, 2006; review), where a greater amplitude P300 occurs in response to invalid events (occurring at temporally unexpected positions). Thus, we favour the idea that temporal predictability eases the cognitive load required for aspects of speech processing, and that this subsequently impacts on motor responses.

We will now consider the P300 latency. In the temporal orienting literature, an earlier P300 peak latency has been found in response to targets which occur at attended locations (Correa et al., 2006; Doherty, Rao, Mesulam, & Nobre, 2005; Griffin, Miniussi, & Nobre, 2002; Miniussi, Wilding, Coull, & Nobre, 1999). Similarly, we observed a tendency for on-beat (attended) targets to elicit a shorter latency P300 than off-beat (unattended) targets. Since the P300 latency is thought to index classification speed (with longer latencies for longer RTs) (Kutas, McCarthy, & Donchin, 1977; Magliero, Bashore, Coles, & Donchin, 1984) and to originate from the linkage between stimulus perception and response (Verleger, Jâkowski, & Wascher, 2005), the longer latency P300 we observed in off-beat conditions may be related to the longer time required for stimulus evaluation.

To conclude, the behavioural and ERP results suggest that attentional orienting can facilitate the phonological processing of pseudowords. The enhanced performance of phoneme detection during this temporal orienting is, through ERP analyses, thought to be due to enhanced cognitive processing via optimisation of temporal prediction.

4.2. *The effect of metrical expectancy on speech processing*

Though we found no behavioural evidence that phonological processing was facilitated in matched metrical conditions (i.e., a Binary Prime followed by a bisyllabic word and a Ternary prime followed by a trisyllabic word), we did observe an interaction between the Prime and Metre factors. However, this interaction was due to the fact that trisyllabic words were processed faster than bisyllabic words regardless of the prime which preceded (i.e., main effect of number of syllables). This was possibly due to the longer 'warning' period before the presentation of the target phoneme, (always at the last syllable); two syllables in trisyllabic words compared to only one syllable in bisyllabic words. Despite this, we did observe ERP correlates to suggest there was indeed an effect of metre on word processing.

The larger N100 amplitude we observed in response to a metrical mismatch is thought to reflect a violation of rhythmical expectations. In simpler terms, metrical expectations which were induced by the prime sequence impacted on the processing of the following word. A similar effect is reported elsewhere in the literature; a larger early negativity in response to 'metrically' unpredictable words (Magne et al., 2007; Rothermich, 2012), though this was found to be more within the 250–450 m s range. Moreover, Rothermich and colleagues also found that an interaction occurred between semantic and metric processing which suggests that a regular metric patterning also facilitates other linguistic processes, via induction of highly predictable temporal expectations.

This larger N100 response to metrical mismatches is also reminiscent of a mismatched negativity (MMN) response elicited from the primary auditory cortex (Sams, Kaukoranta, Hämäläinen, & Näätänen, 1991), although MMN is typically recorded in passive tasks. The MMN response is triggered by deviant sounds in a repetitive sequence (Näätänen, 1990), and reflects the brains ability to compare incoming auditory information against the short-term memory trace of auditory features. This memory trace is reported to be between 6–10 s (Sams, Hari, Rif, & Knuutila, 1993), and thus could form over the short course of our repetitive rhythmical prime for the binary and ternary unit. The MMN might also represent a more general mechanism of expectancy violation which is also sensitive to global target probabilities. Deviants do not have to be deviations of acoustic features (pitch/duration) but can also be violations of more complex, metrical violations (Vuust et al., 2005) or sequential patterns (Herholz, Claudia Lappe, & Pantev, 2009; List, Justus, Robertson, & Bentin, 2007). Indeed, the response we observed seems to suggest that the pseudoword was processed as a continuation of the rhythmic prime which either matched or mismatched the presented metrical structure. As such, this may also be taken as evidence that the processing of metrical features in speech and in the auditory domain in general are similar. In other words there seems to be a common level of representation of temporal structure, possibly defining grouping and temporal hierarchy as suggested by Lerhdal and Jackendoff (1983).

In sum, the larger negative component elicited in response to target words which were metrically mismatched to the prime structure can be linked to the detection of auditory events which do not conform to expectations. Unlike beat expectations discussed above, metrical expectations were not induced through regularity or entrainment to the rhythmic prime, since the pseudoword which followed it was spoken naturally and thus contained syllables of unequal durations. As such, there can be said to be an isomorphism between bisyllabic or trisyllabic pseudowords and binary or ternary prime structure, either in terms of metrical structure (weak and strong elements), or, more simply, in terms of syllable number. In either case, it is clear that a rhythmically 'matching' prime (in terms of either metrical stress or at the least in terms of numeracy), resulted in an N100 correlate which indexes a facilitated processing. The absence of any behavioural evidence for this is not unusual since the ERP method is more able to detect subtle brain processes which may not detected by behavioural measures (Kotz, 2008). In addition, metric processing seems to occur independently of attention (Magne et al., 2007). With a stronger violation of expectations, perhaps through enhancing expectations through the use of a

---

[1] Because 'off-beat' placements were consistently one syllable later than 'on-beat' placements, one might ask whether the effect we report may in fact be due to the temporal delay of the target rather than beat alignment itself. However, if the temporal delay were indeed responsible for the larger P300 amplitude and latency (rather than our interpretation as an effect of off-beat alignment), one would expect target words with similar onset times to elicit a similar P300, such as 'on-beat' bisyllabic words vs. 'off-beat' trisyllabic words (in both, the onset is at 1 syllable before the 'beat'). However we found that even though target words in the two different conditions had similar onsets, the off-beat condition resulted in a significantly higher amplitude P300 component, which well-reflects our interpretation of these results as an effect of beat alignment.

longer prime, the effect of metrical incongruency might also reveal itself in behavioural RT results.

Finally, an interesting point is the lack of interaction between beat and metre factors, possibly pointing to independent processing. While one may expect to find a significant interaction between these two factors, because meter is necessarily reliant on beat, the cognitive mechanisms that we manipulated are orienting of attention in time (on-beat vs. off-beat) and sequencing (matched metre vs. mismatched metre). In this respect it is less surprising to have a lack of interaction, insofar as these mechanisms may rely on different neural resources, left parietal cortex for the former (Coull, Frith, Büchel, & Nobre, 2000) and inferior frontal gyrus for the latter (Vuust, Roepstorff, Wallentin, Mouridsen, Østergaard, 2006).

4.3. Perspectives

We will briefly consider the experimental stimuli used and will then conclude by suggesting how our results may be applicable to therapeutic methods used for speech rehabilitation.

We used pseudowords spoken by a native French speaker. The interest of using pseudowords is that ‘purely’ phonological processing can be investigated without the influences of lexical knowledge (Ganong, 1980; Pitt & Samuel, 1995). Despite the advantages of using such a simplistic design, additional investigations would be required in order to examine to what extent these results are applicable to true speech: having shown here that phonological processing is enhanced, the next steps would be to use real words. Considering, too, that a regular sentence has a more evident rhythm than an isolated word, a further step will be to adapt this priming paradigm for sentences. This would allow for further speculations to be made about how the additional ‘top–down’ influences (e.g., semantic and syntactic) may impact upon phonological processing.

The fact that pseudowords were both spoken and heard by native French speakers is also worth considering. Since French words always contain a final stressed syllable, this prosodic stress was most likely present in the production and perception of pseudoword stimuli. Though this has no bearing on the interpretation of our results (since this means that metrical structure of words was consistently regular), it would be interesting to see whether a language without this regular placement of stressed syllables such as English would be as sensitive to the same priming paradigm. It might be predicted that words without the metrically regular structure of French may be more difficult to prime for since there are a greater number of metrical possibilities (with both trochaic and iambic word stress patterns) to which the listener would be familiar with, and therefore rhythmical expectations may be more varied.

From a broader perspective, our results may be relevant when thinking about current speech therapies. For example, Rhythmic Speech Cuing (Thaut, 2005a) is used for pacing the production of speech by using, in many cases, ‘patterned’ cues which place a beat on salient syllables in speech. That priming for beat position (in the present study) seems to enhance phonological processing supports the validity of this method. In addition, priming for metrical stress may enhance speech acquisition (Jusczyk, 1999). With regards to speech production, a similar method could apply to cochlear implanted children who experience a delayed speech acquisition, especially considering that whilst they have a variable pitch perception, temporal and amplitude (rhythmic) cues are the principle cues conveyed by this prosthesis (Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995). In addition, CI children display an impaired prosody production (Lenden & Flipsen, 2007) which may be alleviated with priming for prosodic stress and rhythm. In addition, research into the mnemonic effects of rhythmical templates has found that entrainment to music can enhance neuronal synchronisation which is associated with improved word recall in patients with Multiple Sclerosis (and in unimpaired populations), thus providing evidence that a regular temporal structure can aid learning and memory (Thaut, Peterson, & McIntosh, 2005b). Priming with a repetitive rhythm matched to a speech rhythm as we have done creates a working memory trace representing a temporal structure which can be helpful for perceiving speech. Finally, based on recent findings that rhythm may benefit speech production in non-fluent aphasics (Stahl, Kotz, Henseler, Turner, & Geyer, 2011), who display deficient intonation and stress patterns, we would predict that rhythmic priming would also enhance speech production in this population.

Acknowledgements

This research was supported by a grant from EBRAMUS ITN (Europe BRAin and MUSic) (Grant Agreement number 238157). We thank Yohana Leveque and Clément François for helping us prepare the stimuli, Nadia Mountaj for help with French translations of experimental material, and Azizuddin Khan who helped to run part of the experiment. We are also grateful to Barbara Tillman, Eleanor Harding, Irene Alonso Fernández, and two anonymous reviewers for their attentive reading of the manuscript and their valuable comments. We also wish to thank all the participants who took part in this study.

Appendix A. Supplementary information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.neuropsychologia.2012.07.018.

References

Barnes, R., & Jones, M. R. (2000). Expectancy, attention, and time. *Cognitive Psychology*, *41*, 254–311.

Boersma, P. & Weenink, D. (2012) *Praat: doing phonetics by computer* [Computer program]. Version 5.3.05, 〈http://www.praat.org/〉(accessed 20 February 2012).

Correa, Á, Lupiáñez, J., Madrid, E., & Tudela, P. (2006). Temporal attention enhances early visual processing: a review and new evidence from event-related potentials. *Brain Research*, *1076*, 116–128.

Coull, J. T., Frith, C. D., Büchel, C., & Nobre, A. C. (2000). Orienting attention in time: behavioural and neuroanatomical distinction between exogenous and endogenous shifts. *Neuropsychologia*, *38*, 808–819.

Cummins, F., & Port, R. (1998). Rhythmic constraints on stress timing in English. *Journal of Phonetics*, *26*, 145–171.

Dilley, L. C., & McAuley, J. D. (2008). Distal prosodic context affects word segmentation and lexical processing. *Journal of Memory and Language*, *59*, 294–311.

Doherty, J. R., Rao, A., Mesulam, M. M., & Nobre, A. C. (2005). Synergistic effect of combined temporal and spatial expectations on visual attention. *The Journal of Neuroscience*, *25*, 8259–8266.

Donchin, E., Karis, D., Bashore, T. R., Coles, M. G. H., & Gratton, G. (1986). Cognitive psychophysiology: systems, processes, and applications. In: M. G.H. Coles, E. Donchin, & S. Porges (Eds.), *Psychophysiology: systems, processes, and applications* (pp. 244–267). New York: The Guilford Press.

Duncan-Johnson, C. C., & Donchin, E. (1977). On quantifying surprise: the variation of event-related potentials with subjective probability. *Psychophysiology*, *14*, 456–467.

Ganong, W. F. (1980). Phonetic categorisation in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, *6*, 110–125.

Griffin, I. C., Miniussi, C., & Nobre, A. (2002). Multiple mechanisms of selective attention: differential modulation of stimulus processing by attention to space or time. *Neuropsychologia*, *40*, 2325–2340.

Herholz, S.C., Claudia Lappe, C. & Pantev, C. (2009). Looking for a pattern: an MEG study on the abstract mismatch negativity in musicians and nonmusicians. BMC Neuroscience 2009, 10.

Hirst, D. J., & Di Cristo, A. (1998). A survey of intonation systems. In: D. J. Hirst, & A. Di Cristo (Eds.), *Intonation systems: a survey of twenty languages* (pp. 1–44). Cambridge: Cambridge University Press.

Jones, M. R., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review*, *96*, 459–491.

Jones, M. R., Boltz, M., & Kidd, G. (1982). Controlled attending as a function of melodic and temporal context. *Perception and Psychophysics*, *32*, 211–238.

Jones, M. R., Moynihan, H., MacKenzie, N., & Puente, J. (2002). Temporal aspects of stimulus-driven attending in dynamic arrays. *Psychological Science*, *13*, 313–319.

Jusczyk, P. (1999). How infants begin to extract words from speech. *Trends in Cognitive Sciences*, *3*, 323–328.

Kotz, S. A. (2008). A critical review of ERP and fMRI evidence on L2 syntactic processing. *Brain and Language*, *109*, 68–74.

Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: P300 as a measure of stimulus evaluation time. *Science*, *197*, 792–795.

Large, E. W., & Palmer, C. (2002). Perceiving temporal regularity in music. *Cognitive Science*, *26*, 1–37.

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: how people track time-varying events. *Psychological Review*, *106*, 119–159.

Large, E. W., & Snyder, J. S. (2009). Pulse and meter as neural resonance.. *The Neurosciences and Music III—Disorders and Plasticity; Annals of the New York Academy of Sciences*, *1169*, 46–57.

Lehiste, I. (1977). Isochrony reconsidered. *Journal of Phonetics*, *5*, 253–263.

Lehiste, I. (1980). Phonetic manifestation of syntactic structure in English. *Annual Bulletin Research Institute of Logopedics and Phoniatrics, University of Tokyo*, *14*, 1–27.

Lenden, J. M., & Flipsen, P. (2007). Prosody and voice characteristics of children with cochlear implants. *Journal of Communication Disorders*, *40*, 66–81.

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge: MA: MIT Press.

List, A., Justus, T., Robertson, L. C., & Bentin, S. (2007). A mismatch negativity study of local–global auditory processing. *Brain Research*, *11*, 122–133.

Magliero, A., Bashore, T. R., Coles, M. G. H., & Donchin, E. (1984). On the dependence of P300 latency on stimulus evaluation processes. *Psychophysiology*, *21*, 171–186.

Magne, C., Astesano, C., Aramaki, M., Ystad, S., Kronland-Martinet, R., & Besson, M. (2007). Influence of syllabic lengthening on semantic processing in spoken French: behavioral and electrophysiological evidence. *Cerebral Cortex*, *17*, 2659–2668.

Makeig, S., Bell, A. J., Jung, T.-P., & Sejnowski, T. (1996). Independent component analysis of electroencephalographic data. In: D. S. Touretzky, M. C. Mozer, & M. E. Hasselmo (Eds.), *Advances in neural information processing systems* (p. 8). Cambridge: MA: MIT Press.

Martin, J. (1972). Rhythmic (hierarchical) versus serial structure in speech and other behavior. *Psychological Review*, *79*, 487–509.

Meltzer, R. H., Martin, J. G., Mills, C. B., Imhoff, D. L., & Zohar, D. (1976). Reaction time to temporally-displaced phoneme targets in continuous speech. *Journal of Experimental Psychology: Human Perception and Performance*, *2*, 277–290.

Miniussi, C., Wilding, E. L., Coull, J. T., & Nobre, A. C. (1999). Orienting attention in time. *Brain*, *122*, 1507–1518.

Näätänen, R. (1990). The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. *Behavioral Brain Sciences*, *13*, 201–288.

New, B., Pallier, C., Ferrand, L., & Matos, R. (2001). Une base de données lexicales du français contemporain sur internet: LEXIQUE. *L'Année Psychologique*, *101*, 447–462.

Patel, A. (2008). *Music, language and the brain*. New York: Oxford University Press.

Pitt, M. A., & Samuel, A. G. (1990). The use of rhythm in attending to speech. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 564–573.

Pitt, M. A., & Samuel, A. G. (1995). Lexical and sublexical feedback in auditory word recognition. *Cognitive Psychology*, *29*, 149–188.

Port, R. F. (2003). Meter and speech. *Journal of Phonetics*, *31*, 599–611.

Quené, H., & Port, R. F. (2005). Effects of timing regularity and metrical expectancy on spoken-word perception. *Phonetica*, *62*, 1–13.

Rothermich, Schmidt-Kassow, & Kotz (2012). Rhythm's gonna get you: regular meter facilitates semantic sentence processing. *Neuropsychologia*, *50*, 232–244.

Sams, M., Hari, R., Rif, J., & Knuutila, J. (1993). The human auditory sensory memory trace persists about 10 s: neuromagnetic evidence. *Journal of Cognitive Neuroscience*, *5*, 363–370.

Sams, M., Kaukoranta, E., Hämäläinen, M., & Näätänen, R. (1991). Cortical activity elicited by changes in auditory stimuli: different sources for the magnetic N100m and mismatch responses. *Psychophysiology*, *28*, 21–29.

Schmidt-Kassow, M., & Kotz, S. A. (2009). Attention and perceptual regularity in speech. *NeuroReport*, *20*, 1643–1647.

Shannon, R. V., Zeng, F., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, *270*, 303–304.

Shields, J. L., McHugh, A., & Martin, J. G. (1974). Reaction time to phoneme targets as a function of rhythmic cues in continuous speech. *Journal of Experimental Psychology*, *102*, 250–255.

Squires, K., Wickens, C., Squires, N. K., & Donchin, E. (1976). The effect of stimulus sequence on the waveform of the cortical event-related potential. *Science*, *193*, 1142–1146.

Stahl, B., Kotz, S. A., Henseler, I., Turner, R., & Geyer, S. (2011). Rhythm in disguise: why singing may not hold the key to recovery from aphasia. *Brain*, *134*, 3083–3093.

Thaut, M. H. (2005a). Neurologic music therapy in speech and language rehabilitation. In: M. H. Thaut (Ed.), *Rhythm, music, and the brain: scientific foundations and clinical applications* (pp. 165–178). New York: Routledge.

Thaut, M. H., Kenyon, G. P., Schauer, M. L., & McIntosh, G. C. (1999). The connection between rhythmicity and brain function. *IEEE Engineering in Medicine and Biology*, *18*, 101–108.

Thaut, M., Peterson, D., & McIntosh, G. (2005b). Temporal entrainment of cognitive functions: musical mnemonics induce brain plasticity and oscillatory synchrony in neural networks underlying memory. *The Neurosciences and Music II: From Perception to Performance*, *1060*, 243–254.

Trehub, S. E. (2001). Musical predispositions in infancy. *Annals of the New York Academy of Sciences*, *930*, 1–16.

Tueting, P., Sutton, S., & Zubin, J. (1970). Quantitative evoked potential correlates of the probability of events. *Psychophysiology*, *7*, 385–394.

Verleger, R., Jâkowski, P., & Wascher, E. (2005). Evidence for an integrative role of P3b in linking reaction to perception. *Journal of Psychophysiology*, *19* 150–150.

Vuust, P., Pallesen, K. J., Bailey, C., van Zuijen, T. L., Gjedde, A., Roepstorff, A., et al. (2005). To musicians, the message is in the meter. Pre-attentive neuronal responses to incongruent rhythm are left-lateralized in musicians. *NeuroImage*, *24*, 560–564.

Vuust, P., Roepstorff, A., Wallentin, M., Mouridsen, K., & Østergaard, L. (2006). It don't mean a thing … :keeping the rhythm during polyrhythmic tension, activates language areas (BA47). *Neuroimage*, *31*, 832–841.

EXPERIMENT 2

Title: Bridging music and speech rhythm: rhythmic priming and audio-motor training effect speech perception

Author names and affiliations: Nia Cason[a,*], Corine Astésano[b,c] & Daniele Schön[a].

a) U 1106, Institut de neurosciences des systèmes, AMU, Centre MEG, Service de Neurophysiologie Clinique, CHU Timone, Rue Saint Pierre 13385, Marseille, France

b) UMR 7309, Laboratoire Parole et Langage, CNRS & Aix-Marseille University, 5 avenue Pasteur, 13006 Aix-en-Provence, corine.astesano@lpl-aix.fr

c) EA 4156, U.R.I. Octogone-Lordat, 5 allées Antonio Machado, 31058 Toulouse Cedex 09, France, corine.astesano@univ-tlse2.fr

* Corresponding author. Email Address: nia.cason@etu.univ-amu.fr (N. Cason)

Contact Details:

Nia Cason: nia.cason@etu.univ-amu.fr, Tel: 0033 (0) 4 91 38 55 77

Corine Astésano: corine.astesano@univ-tlse2.fr, Tel: 0033 (0) 5 61 50 24 69

Daniele Schön: daniele.schon@univ-amu.fr Tel: 0033 (0) 4 91 38 55 77

Abstract

We investigated whether complex metrical rhythms corresponding to the prosodic structure of sentences can enhance phonological processing via the induction of rhythmical expectations (priming). We also investigated to what extent this priming effect is sensitive to rhythmic audio-motor training.

Participants listened to a metrical prime followed by a sentence and performed a phoneme detection task. The prime's meter either matched or mismatched the prosody of the sentence in terms of both stress patterns and number of elements. Behavioural (reaction time (RT)) data were collected from two groups of speech-fluent listeners: a group who underwent audio-motor training with the musical rhythms heard in the experimental task (AM group), and a group who did not receive this audio-motor training and whose only exposure to the musical rhythms was in the auditory modality (A group) in the experimental task. We hypothesised that 1) metrical priming (inducing metrical expectations about sentence prosody through a musical rhythmic sequence) would enhance the phonological processing of sentences when the prime rhythm and sentence rhythm were matched, and 2) that audio-motor training with musical rhythms would further enhance this priming effect. Providing a matching rhythmic prime context resulted in faster RTs, thus revealing a cross-domain effect of musical rhythm on the phonological processing of speech. In addition, there were differences between the two groups which are suggestive of a greater priming effect with sensorimotor engagement with musical rhythm. These results may have important implications about how rhythm is used in speech therapies, and suggest that rhythm in music and speech may rely, at least to a certain extent, on shared temporal processing brain resources.

Keywords: Speech; Music, Prosody, Meter, Rhythm

Abbreviations: A group (Auditory modality exposure group); AM group (Audio-motor training group); Reaction Times (RTs).

1. Introduction

Though speech and music may seem distinct from one another, both are hierarchically-organised, rule-based systems for which our perceptual systems are primed (Trehub, 2001). The similarities between music and speech processing have been subject of much current interest (Jäncke, 2012; Patel, 2011). Here, the current focus is ‘rhythm’, and how cross-domain similarities can explain the potential for musical rhythm to impact on speech processing. Rhythm can be broadly defined as the temporal organisation of acoustic events which unfold over time. It often gives rise to a sense of pulse or beat, a series of regular and recurrent psychological events (Cooper & Meyer, 1960). As well as ‘rhythm’, both speech and music have ‘meter’, which can be described as an emergent temporal structure resulting in a hierarchical organisation of beats (e.g. 'strong’ and ‘weak’).

Western music typically has a binary, march-like metre (1**2** 1**2** 1**2**) or a ternary, waltz-like metre (12**3** 12**3** 12**3**). The emergence of metre from a sequence of rhythmic sounds is typically automatic (London, 2012), and the ability for humans to distinguish binary and ternary metrical patterns is present from infanthood (Hannon & Johnson, 2005). Even on hearing an isochronous sequence of identical sounds, humans automatically project a metrical structure onto them, engendering a perceptual illusion of weak and strong elements (Bolton, 1894). Recognition of metrical patterns can also occur implicitly, without requiring conscious attention (Schultz, Stevens, Keller &Tillman, 2013).

Similarly, strong and weak elements (syllables) form the metrical patterning of utterances. Though speech does not possess the same degree of regularity as music (Patel, 2008), metrical ‘rules’ may allow speech rhythm to be made predictable; for example, a final stressed syllable marks the end of word groups in French (Hirst & Di Cristo, 1998). There are also different types of prosodic stress (e.g. lexical stress, pitch stress, emphatic stress) which have different functions in speech, and different degrees of stress (e.g. primary and secondary lexical stress; high, low or complex pitch stress). Much like in music, it is the interactions between these multiple levels of stress in speech which allow for the emergence of speech rhythm (Handel, 1989: pp. 383), which can in turn facilitate on-line speech perception in a context-dependent, long-distant manner (Dilley & McAuley, 2008).

In both music and speech, metre allows a listener to form predictions about what will happen next. For instance, on hearing a ternary metrical sequence (12**3** 12**3** 12**3**), we are able to automatically predict that a ‘12**3**’ (i.e. weak-weak-strong pattern) will follow. Similarly, if a listener hears a list of trisyllabic words with a final stress, they will expect a word with the same structure to follow (Pitt & Samuel, 1990). The function of metrical contrast seems to be attentional in both speech (Pitt & Samuel, 1990) and in music (Jones, 2008; Large and Palmer, 2002). For instance, The Attentional Bounce hypothesis (Pitt & Samuel, 1990) states that attention is oriented to syllables which are expected to be stressed. It claims that the position of these stressed syllables can be predicted on the bases of the metrical patterns of speech, and that this is reflected by quicker phoneme detection at attended syllables. In music, the Dynamic Attending Theory predicts that on hearing a metrical rhythm, there is a coupling between external and internal oscillators such that there is a dynamical allocation of attention to strong and weak elements, with more attention being allocated at strong metrical positions (Jones, 2008; Large and Jones, 1999; London, 2012; Snyder and Large, 2005).

Considering that 1) metre exists in both music and speech and that 2) similar accounts exist for the psychological impact of metrical rhythm in both speech and music, it may be that metre perception in these two domains is dependent on similar attentional mechanisms (Cason & Schön, 2012). It can therefore be predicted that expectations induced by musical metre may impact on the processing of speech metre, and that this will consequently impact on phonological processing of speech.

Inducing rhythmic expectations can enhance auditory processing in both the speech domain and the music domain. In the speech domain, presentation of metrically identical words induces metrical expectations which enhance the processing of a target word when it shares the same metrical pattern (Pitt & Samuel, 1990; Rothermich, Schmidt-Kassow & Kotz., 2011). In the music domain, auditory events occurring at points at which the beat is expected to occur are better processed (Barnes & Jones, 2000; Ellis & Jones, 2010; Jones, Boltz & Kidd, 1982; Jones, Moynihan, MacKenzie & Puente, 2002). The cross-domain effect of musical rhythm on speech processing has also been investigated (Gordon, Magne & Large, 2011), where it was found that temporal alignment between stressed beats and stressed syllables allowed for a greater comprehension of speech. In a recent experiment (Cason & Schön, 2012), we investigated whether inducing rhythmic expectations can have also a cross-domain effect on subsequent speech processing. Expectations about beat and metre were induced by a musical rhythmic prime, which was followed by a bi- or trisyllabic pseudoword. Focussing here on the meter aspect, we found electrophysiological (EEG) evidence of an enhanced processing when the speech metre matched the rhythmic prime metre (for example, a ternary metre (123 123 123) followed by a trisyllabic word with a 'weak-weak-strong' structure). Thus, it seems that musical rhythm impacts on speech processing by inducing domain-general metrical expectations. To build upon these findings, we wanted to see whether this effect, seen at the level of single pseudowords, would also be present for real sentences.

In the experiment described above, both rhythmic primes and speech stimuli were presented auditorily. However, actively engaging in auditory rhythms may further enhance this cross-domain effect by allowing listeners to internalise musical rhythms which are shared by speech. From a clinical standpoint, the importance of audio-motor training with musical rhythm in enhancing speech fluency has not been fully acknowledged. In therapies such as Rhythmic Speech Cueing (Thaut, 2005), for example, it is often the therapist who produces the musical rhythm on a drum whilst the patient is asked only to produce speech. In this case, patients only engage auditorily with the musical rhythm and so any further benefits of motor rhythmic engagement are not fully potentiated.

Both speech and music require audio-motor interactions in order to consolidate accurate production and to continually monitor output (Guenther, 2006; Lappe, Herholz, Trainor & Pantev, 2008; Levelt et al., 1999; Zatorre, Chen & Penhune, 2007;). As well as the importance for production, the effect of audio-motor interactions on musical rhythm perception has also been noted. For instance, movement to an ambiguous metrical sequence influences how it is perceived in both infants (Phillips-Silver & Trainor, 2005) and in adults (Phillips-Silver & Trainor, 2007). Motor engagement with musical rhythm also strengthens beat extraction in both musicians and non-musicians compared to listening alone (Su & Pöppel, 2012), and musical training enhances sensitivity to hearing metrical patterns and hierarchies (Geiser, Sandmann, Jancke & Meyer, 2010). There is therefore a strong effect of motor training on the perception of musical rhythm. Considering this, it can be hypothesised that audio-motor training with a musical rhythm increases listeners' sensitivity to its structure, which in turn allows musical rhythm to be more effective in enhancing speech perception.

In the present experiment, a musical metrical prime sequence was used to induce metrical expectations about both stress patterns and the number of elements. The subsequent presentation of a sentence either conformed to the stress pattern/numeracy of the previously presented prime - and thus to listeners' metrical expectations - or did not. We hypothesised that when metrical expectations were met (when the sentence metre was identical to the prime metre), phonological processing of the sentence would be facilitated. A phoneme detection task was employed to measure phonological processing, a building block of speech perception, with faster reaction times (RTs) indicating a facilitated access to phonological information. In addition to this, we tested two groups: one with audio-motor training with the musical rhythms presented in the experiment (Audiomotor (AM) Group), and one without this training (Auditory only (A) Group). We hypothesised that the positive effect of metrical priming on phonological processing would be more striking in the AM Group due to the consolidation of metrical representations through audio-motor training.

2. Methods

2.1 Stimuli

Experimental trials consisted of a rhythmic prime sequence followed by a sentence. We used four rhythmic prime sequences (Prime 1, Prime 2, Prime 3, Prime 4) which comprised 4 experimental blocks, and four sentence types built around these same rhythms (Table 1).

Prime sequences were created in Adobe Audition and differed in number of beats (6 or 7) and in the placement of stress. These rhythmic patterns were comprised of percussion sounds which had a stimulus onset asynchrony of 225 msecs. Initial stressed percussion sounds had a rim shot timbre, a duration of 196 ms and an average Root Mean Square (RMS) of -41.25 dB, final stressed percussion sounds had a snare timbre, a duration of 353 ms and an RMS of -28.88 dB, and weak percussion sounds had a closed high hat timbre, a duration of 138 ms and an RMS of -43.63 dB.

Sentence stimuli were constructed around the four rhythms used for the rhythmic primes; 3 to 4 syllable groups such as those found in these sentence stimuli are the most common meter in French (whatever the speaking style) (Astésano, 2001). The final word of the sentence (in which the target was present 50% of the time) was selected using Lexique 3 (New, Pallier, Ferrand and Matos, 2001). The criteria for these final words were as follows: they were bisyllabic words with a CV/CV structure (at the phonological level), words were balanced for lexical frequency across the four sentence rhythm types, and uncommon words were discarded. The last syllable (open CV syllable) of each word contained a target vowel, as /i/ in /mari/ for 'mari', and the previous consonant macro-class (L= liquid; N = nasal; UV-F: unvoiced Fricative: UV-S: unvoiced stop; V-F: voiced fricative; V-S: voiced stop.) were balanced over the 4 rhythmic conditions.

Sentences were recorded in a soundproof booth by a native French speaker and were spoken within a carrier sentence in order to control for sentence-ending prosodic effects. For instance, from the carrier sentence '"J'ai bu un bon cafe" n'est pas facile a dire' ['"I drank a good coffee" is not easy to say'], we extracted the first part of the sentence 'J'ai bu un bon cafe'. The last part of the carrier sentence had the same prosodic rhythm as the first (extracted) part of the sentence (in the example sentence provided, both parts of the sentence have a "x x x X x X" structure, where "X" denotes a strong syllable). For each target sentence extracted, the duration of the target sentence, the onset of the final word relative to the beginning of the sentence, and the duration of the final syllable were calculated. This was done by determining the syllable onsets and offsets (PRAAT speech segmentation software, Boersma & Weenink, 2012). These duration values were used to determine stimuli presentation and also to timelock RT data to the onset of the final (target) vowel. Audio examples of the four primes and their sentence counterparts are available as Supplementary Material.

2.2 Participants

Behavioural (RT) data were collected from two groups of healthy participants. The 'auditory only' group (A Group) was exposed to the rhythmic primes auditorily during the experimental task, whilst the 'audio-motor' group (AM Group) underwent an additional period of audio-motor training with the rhythmic primes. In the A Group, 17 participants (7 female) between the ages of 22 and 39 years (mean = 28 years, 4 months) participated in the experiment. In the AM Group, 17 participants between the ages of 23 and 42 (mean = 30 years) participated in the experiment. All participants had normal hearing and French was their first language. 4 participants in each group considered themselves to be amateur musicians (AM group: drummer, rapper, percussionist, guitarist/singer; A Group: singer, pianist, drummer/pianist, pianist/accordionist), but there were no professional musicians in either group. Behavioural data were collected during a reaction-time task (phoneme detection task). Each participant gave informed consent prior to experimentation and was compensated for their time with a gift.

2.3 Procedure

For the A Group, participants sat in front of a computer screen. The volume for auditory stimuli was adjusted to a comfortable level. Participants were asked to decide whether the sound associated with the visually-presented target vowel presented on the screen in front of them was or was not present in the final word of the auditorily-presented sentence, and to make their responses as quickly as possible by pressing either a 'yes' button or a 'no' button. The hand in which each button was held was balanced across subjects.

After a training session of 12 trials (3 trials for each of the prime rhythms), the experiment began, consisting of four blocks of 40 trials each. Within each block we used always the same prime (P1, P2, P3 & P4 blocks, order counterbalanced across subjects). For each trial, the target vowel (/a/, /e/, /i/, /y/, /u/ or /o/) was presented on a computer screen (in lowercase) for the duration of the trial, and was present within the heard sentence 50% of the time. In each trial, the prime was followed by a sentence whose first stressed syllable landed 562.5 ms after the final stressed beat of the prime (2.5 x the interval between elements, in order to control for possible entrainment effects). Sentences could either match or mismatch the metrical structure of the prime (in terms of the number of elements and/or the stress pattern). The experimental conditions were as follows: Condition 1) stress and number match (complete match); Condition 2) stress mismatch, number match; Condition 3) partial stress match (first half of the sentence matched the first half of the prime), number mismatch; Condition 4) stress mismatch, number mismatch (full mismatch) (see Table 2 for an example). The order of experimental conditions was pseudo-randomised within each experimental block. Participants heard each sentence only once. However, sentences were used in all four conditions across participants (fully balanced). Error feedback was received at the end of each experimental block. Both visual target phoneme and auditory stimuli were presented using Presentation Software.

For the AM Group, this experimental protocol was identical except for additional auditory-motor training. For this additional training, participants heard the prime rhythm and were asked to copy the rhythm using their mouth; they were asked to distinguish between strong and weak elements of the rhythmic prime by using different sounds (‘ba’ and ‘ka’). For each experimental block (i.e., each prime), participants underwent two periods of this auditory-motor training, once before the experimental block, and once midway through. These training periods consisted of ten repetitions, making a total of twenty audio-motor rehearsals per prime. Repetitions of the prime were recorded using a M-Audio Microtrack 24/95 recorder.

2.4 Data Acquisition and Analyses

Analyses on behavioural RT data were performed for correct trials only. Repeated-measures analysis of variance (RM-ANOVA) was used, with the categorical predictor ‘Group’ as a between-subjects factor (2 levels, A Group or AM Group). Factors included ‘Prime’ (4 levels, Prime 1, Prime 2, Prime 3 or Prime 4), ‘Condition’ (4 levels, Condition1/2/3/4). All p-values reported below were adjusted using the Greenhouse-Geisser correction for nonsphericity, when appropriate, and Fishers tests were used in post-hoc comparisons. Prime reproductions were scored as follows: 1 for a correct reproduction (the stress/number pattern of the prime sequence was reproduced without errors), 0.5 for a correct reproduction but which had a large hesitation, and 0 for incorrect prime reproduction. Of the 17 participants in the AM Group, we collected recordings from 13 participants. The scores for the first five repetitions of each prime were taken. Friedman’s test was used for this reproduction analyses.

Prime Rhythm	Sentence Rhythm
Prime 1 x x x X x X	Le scandal**eux** sén**at** x x x X x X
Prime 2 x x X x x X	Le carr**osse** du coch**er** x x X x x X
Prime 3 x x X x x x X	Chorégraph**ier** le ball**et** x x x X x x X
Prime 4 x x x X x x X	Il prescr**it** le bon cach**et** x x X x x x X

Table 1. Graphic representation of the four rhythms used for experimental stimuli. Sentences were constructed to share the rhythmical structure of the primes.

Condition	
1. Stress M Number M	/a/ → 225 msec x x x X x X → Le scandal**eux** sén**at** (x x x X x X)
2. Stress MM Number M	x x x X x X → Le carr**osse** du coch**er** (x x X x x X)
3. Stress partial M Number MM	x x x X x X → Chorégraph**ier** le ball**et** (x x x X x x X)
4. Stress MM Number MM	x x x X x X → Il prescr**it** le bon cach**et** (x x X x x x X)

Table 2. The four experimental conditions. Primes were followed by a sentence rhythm which either matched both in terms of stress and number (Condition 1), which mismatched the stress pattern (Condition 2), which mismatched the number but partially matched stress (Condition 3), or which

completely mismatched in terms of both stress and number (Condition 4). The examples presented here are of trials from Prime 1 block. A target vowel was presented on-screen. Participants then heard the rhythmic prime followed by a sentence and had to decide whether the vowel target was present in the final syllable of the word.

3. Results

All participants were able to perform the task easily (mean percentage of incorrect/missed trials in the experimental blocks = 5.88% for Auditory-only group (A group), 5.55% for auditory-motor group (AM group)).

A three-way ANOVA was run on correct response data with ‘Group’ (A group/AM group) as a between-subjects factor and ‘Prime’ (Prime 1/Prime 2/Prime 3/Prime 4) and condition (Condition1/2/3/4) as within-subject factors. Additional planned comparison analyses were run on a subset of the data: for the two groups, RTs from Condition 1 (matching metre) were compared against those from Condition 2 / 3 / 4 (mismatching metres).

3.1 The Effect of a Matching Metrical Prime

There was a main effect of Condition (F(3, 96)=3.5072, p=.0182). A post-hoc analysis revealed this to be the result of quicker RTs in Condition 1 (metric match, mean 521.805 msec) compared to those in Condition 2 (p=0.004, mean 547.08 msec), Condition 3 (p=0.064, mean 538.157) and Condition 4 (p=0.0077, mean 545.605) (metrically mismatching conditions).

There was also a Prime x Condition interaction (F(9, 288)=2.342, p=.014), revealing that the main effect of Condition was driven by Prime 2 and Prime 3 (Figure 1): a post-hoc analysis revealed there to be no experimental effect of Condition in either Prime 1 or Prime 4. For Prime 2, condition 1 (matching) was significantly different to condition 2 (p=0.00004), and condition 4 (p=0.0036), and not significantly different to condition 3 (p=0.618). In Prime 3, condition 1 (matching) tended to be different to condition 2 (p=0.077) and was significantly different to condition 3 (p=0.04) and condition 4 (p=0.024).

3.2 The Effect of Prime 2 and Prime 3: A Group and AM Group

Planned comparisons revealed that this effect of Condition seen for Primes 2 and 3 differed between the A and AM Groups. For Prime 2, Condition 1 resulted in faster RTs in both groups (Figure 2a). For the AM Group this was significant (F(3,30)=5.922, p=0.0026) and in the A Group there was a trend

(F(3,30)=2.504, p=0.078). For Prime 3 however, Condition 1 resulted in significantly faster RTs in the AM Group only (F(3,30)=5.048, p=0.005) and was not significant in the A Group (F(3,30)=0.118, p=0.94) (Figure 2b).

3.3 Recording Results during Audio-motor rhythm training

Across the participants in the AM Group, Prime 2 and 3 were also the easiest to reproduce (mean error rates: Prime 1: 9.556%, s.d. 24.09; Prime 2: 1.875 %, s.d. 4.411; Prime 3: 6.736%, s.d. 11.27; Prime 4: 10.417%, s.d. 13.97).

Taking the first 10 repetitions of each prime (during which learning occurred), Prime 2 repetitions were less variable and almost at ceiling performance. Prime 1 and Prime 4 were reproduced with the least accuracy and more variability. This effect did not reach statistical significance but did show a trend nonetheless (χ^2(12, 3) = 6.3, p = 0.09789). In the remainder of trials (repetitions 11 – 20), there was no significant difference between prime reproduction scores (χ^2(12, 3) = 1.8, p = .614), with mean error rates at: Prime 1: 11.055%, s.d. 31.32; Prime 2: 0.4516%, s.d. 1.44; Prime 3: 2.38%, s.d. 4.2; Prime 4: 7.5%, s.d. 16.98.

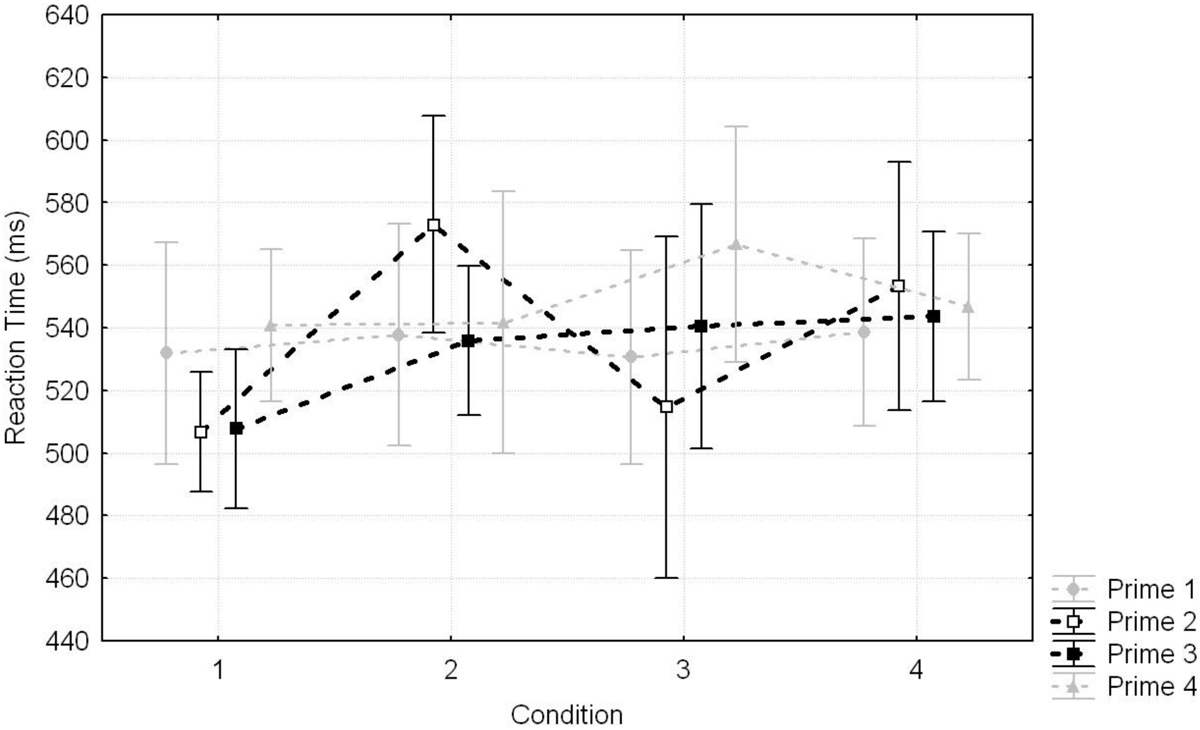


Fig. 1. Whilst Prime 1 and Prime 4 yielded no significant differences between Conditions, Condition 1 (matching condition) in Prime 2 and Prime 3 resulted in faster RTs compared to mismatching Conditions.

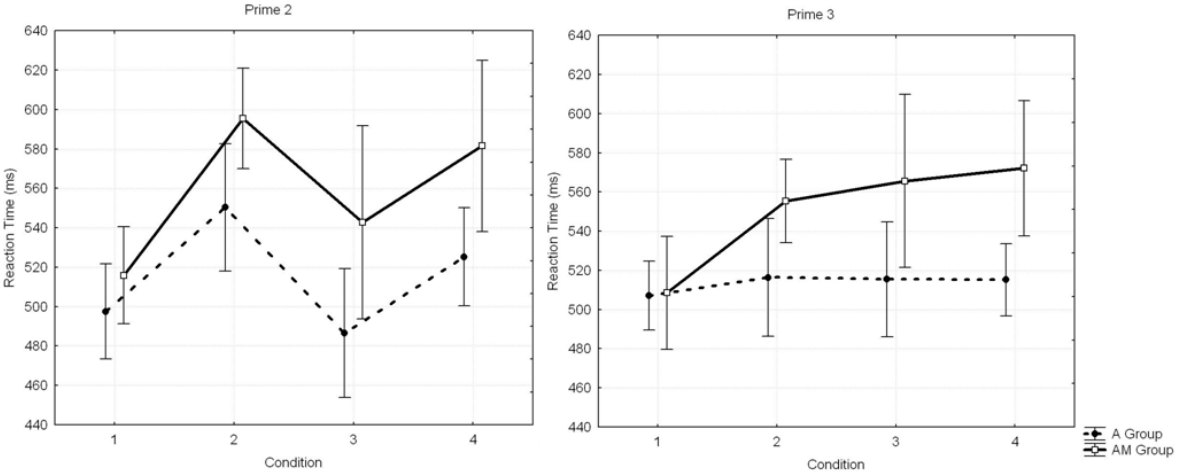


Fig. 2. ***2a, left:*** Within Prime 2, Condition 1 (matching condition) resulted in faster RTs in both groups. This difference was more pronounced for the AM Group (bold line). ***2b, right:*** Within Prime 3, Condition 1 resulted in faster RTs in the AM Group only.

4. Discussion

The aim of this experiment was two-fold. First, we wanted to extend previous findings that musically-induced rhythmic expectations (rhythmic priming) can impact on speech processing of spoken pseudowords (Cason & Schön, 2012). We did this by testing whether rhythmic primes would allow a listener to form prosodic expectations about a subsequently presented sentence. Second, we wanted to investigate how additional audio-motor training with musical rhythms influences this cross-domain effect.

We found that providing listeners with a matching metrical prime resulted in quicker RTs to a phoneme detection task. This effect was more prominent with primes that could be easily reproduced and when participants were trained to reproduce the prime before the experiment. We will first discuss the effect of rhythmic priming on speech processing and will then consider how additional audio-motor training with musical rhythm may enhance sensitivity to speech rhythms.

4.1 The effect of metrical priming on sentence processing

In the experimental task, the prosodic structure of heard sentences either matched or mismatched a prior metrical prime in terms of both stress patterning and/or number of elements. RTs were taken as a measure of sentence processing, with quicker RTs indicating a facilitated access to phonological information. Here, we will discuss the finding that prosodic predictability, induced through rhythmic priming, can enhance sentence processing.

The number of prime sounds was mirrored by the number of syllables in the sentence (Condition 1 and 2) or not (Condition 3 and 4). In the case of stress patterning, the weak and strong prime sounds were fully mirrored by the sentence prosody (Condition 1), partially mirrored by the sentence prosody (Condition 3) or were not mirrored by the sentence prosody (Condition 2 and 4). In partially matching stress patterning (Condition 3), the first half of the sentence conformed to listeners' expectations and diverged from prosodic expectations at the 6th syllable whilst in fully mismatching stress patterning (Conditions 2 and 4), the first stressed syllable occurred either one syllable earlier or one syllable later than anticipated. These four conditions were chosen in order to give an indication about which aspects of a rhythmic cue (stress pattern, number of elements, or both) might be stronger for inducing prosodic expectations. We found that a rhythmic pattern which well-primed for the target sentence (Condition 1) resulted in an enhanced phonological processing of sentences; RTs were significantly quicker than stress mismatching conditions (Conditions 2 and 4) and were almost significantly quicker than the partial stress match condition (Condition 3). The fact that a partially matching stress context resulted in RTs which were more similar to metrically matching conditions could indicate that stress patterning was a more salient cue than numerical context. Overall, it seems that providing a listener with a rhythmic context, matching the target sentence in terms of both number and stress patterning, results in an enhanced sentence processing.

The findings that rhythmic predictability can enhance auditory processing is central to the hypotheses presented in this paper. Musical rhythm may induce expectations about 'when' events are expected to occur (Jones & Boltz 1989; Large & Jones, 1999), as well as 'what' events are expected to occur. For instance, on hearing a musical rhythm a listener forms expectations about when in time a future event will occur, and events conforming to these expectations are better perceived in both music (Jones et al., 1982) as well as speech (Cason & Schön, 2012). The current interest was the ability for rhythmic priming to induce cross-domain expectations about 'what' speech events will occur: for instance whether hearing a weak-weak-strong musical rhythmic structure can induce expectations for a weak-weak-strong prosodic structure in speech. We reasoned that hearing musical metrical patterns would have the ability to induce expectations about speech prosody, and in isolation from beat expectations since the formation of prosodic expectations do not require temporally isochronous stimuli (Pitt & Samuel, 1990). However, in a study by Quené and Port (2005), the authors found that whilst temporal regularity (priming for 'beat') had the ability to enhance phonological processing of a target word, inducing metrical expectations did not. This study differs from our own in several ways. Firstly, their target stimuli were single words whilst our own were sentences. In addition, we purposefully obliterated any effect of entrainment by positioning the first (stressed) syllable off-beat, thus we measured only the effects of metrical expectancies on sentence processing. Lastly, rather than using speech itself to induce metrical expectations, we used a musical metrical prime. There are several reasons as to why musical rhythm may hold a special ability to induce prosodic expectations over that of speech itself. Most would agree that the key features of music are rhythm and metre, because they directly deserve one of the most important functions of music: the ability to synchronise. While synchronisation is not afforded to the same degree by natural speech, temporal features of speech are nonetheless vital for perception and comprehension: when these prosodic features are clearly primed by music they may affect the different levels of speech processing more strongly that regular speech alone. These different levels may be at the phonological level (Cason & Schön, 2012; Pitt & Samuel, 1990; Quene & Port, 2005), syntactic level (Roncaglia-Denissen, Schmidt-Kassow & Kotz, 2013; Schmidt-Kassow & Kotz, 2009) and also semantic level of processing (Rothermich et al., 2011). We would therefore expect the results found here for phonological processing to apply to these other levels of speech processing, too.

4.2 A differential effect of the four rhythmic primes

There was a difference between primes in their ability to impact on sentence processing: Prime 2 (a “x x X x x X” pattern) and Prime 3 (a “x x X x x x X” pattern) drove the main effect of rhythmic priming on phonological processing; matching conditions (Condition 1) resulted in quicker RTs for these two primes only. Interestingly, whilst this was true in both groups for Prime 2, this effect was only apparent for Prime 3 in the AM group. The question arises as to why one particular rhythmic sequence was more ‘accessible’ to both groups, whilst in the case of Prime 3, it required an additional audio-motor training for an effect to be observed. We will now discuss why these two rhythms in particular may have been most successful in priming for speech prosody, and will then speculate as to how additional audio-motor training may have facilitated the effect of Prime 3.

Since a priming effect of Prime 2 was present for both groups, we can say that this rhythm in particular was most successful in inducing prosodic expectations. There are two possibilities as to why. First, it is the only prime with a repeating structure, whilst the other primes involved no repetition of the same unit. Words used in the current experiment were most often two-syllable words, hence, when preceded by an article, this most commonly resulted in a weak-weak-strong syllable metre (as French typically has a final accent). In this case, the representation of metre, chunked into two groups of three, may have become stronger. Gestalt principles pose that regularity and structure characterise motor rhythms, meaning that in speech motor rhythms, the repetition of the same structure is favoured (Pasdeloup, 2005). This was also apparent in the audio recordings taken from the AM Group during the audio-motor training phase: Prime 2 was also the easiest to reproduce compared the other three primes. Because the four primes were similar in terms of number of sounds, sound durations and tempo, this is thought to further reflect the presence of a more solidified representation of its rhythmic structure and thus shows an interesting link between perception and production: the Prime 2 rhythm, more easily reproduced by participants, also exerted the greatest perceptual priming effect.

Second, there is the possibility that the metric structure of Prime 2 is more representative of speech rhythms typically found in spoken French. We built sentences around four template rhythms constructed solely for the purposes of this experiment, and which may have therefore been differentially representative of French speech rhythms. In this case, listeners’ native language (French) may have played a role in the ability for the rhythmical prime to influence perception. Indeed, the characteristic rhythms of one’s native language can influence musical rhythm perception (Iversen, Patel & Ohgushi, 2008; Kusumoto & Moreton, 1997). To summarise this point, a more ‘familiar’ sentence rhythm may have been more prone to disruption in mismatching conditions and/or more prone to facilitating effects in matching conditions. We will next discuss the finding that Prime 3 only exerted a priming effect within the AM Group.

4.3 The effect of audio-motor training on cross-domain speech processing

Whilst both groups completed the same experimental task, the AM Group underwent an additional period of audio-motor training with the musical rhythms used in the experiment (prior to and during the experiment). Over the four Primes, audio-motor training with the musical rhythm did not enhance the priming effect of musical rhythm, although it did result in an enhanced priming effect for Prime 3. This may denote an increased sensitivity to musical rhythm through audio-motor training.

Whilst the effect of priming between the two groups did not reach significance, further analyses revealed that whilst Prime 3 resulted in no significant differences between priming conditions in the A Group, it was an effective prime within the AM Group: matching metre conditions (Condition 1) resulted in quicker RTs to the phoneme detection task than did mismatching metre conditions. Important to note, however, is that the metrically matching condition (Condition 1) in Prime 3 did not seem to enhance processing in the AM Group; rather, mismatching metre conditions (Conditions 2, 3 & 4) resulted in slower RTs. That is to say, primes with an incongruent metrical pattern to the target sentence prosody resulted in a more delayed processing for listeners in the AM Group. Nonetheless, this evidences a cross-domain effect of the audio-motor training on sentence perception, perhaps through a reinforcement of metrical representations and consequent increase in sensitivity to metrical patterns. This enhanced sensitivity to the Prime 3 rhythm could result in a greater conflict in mismatching conditions.

5. Conclusions

We have replicated previous findings that musical rhythm can induce implicit prosodic expectations, and that this impacts upon the processing of subsequent speech (Cason & Schön, 2012). We have shown this to be the case for real speech (as opposed to pseudowords) and in sentence contexts (as opposed to single words). Since no participants claimed to be aware of the rhythmic relationship between the primes and sentences, these results also provide evidence that the metre of speech was implicitly recognised, much like that of music can be (Schultz et al., 2013), and that listeners do not need to be aware of this relationship for it to impact on their speech processing. Moreover, the findings that there is a differential effect of the rhythmical Primes used in the experiment suggests that linguistic rhythmic familiarity may play a role in the priming effect.

Interestingly, audio-motor training seems to further sensitise listeners to musical rhythm which in turn increases the potential for musical rhythm to exert an effect on sentence perception. The ability for audio-motor training to amplify the cross-domain effect from music to speech in this way might lend support to the usefulness of musical rhythm in speech rehabilitation contexts. However, the fact that mismatching conditions within Prime 3 seemed to result in an inhibition of phonological processing (rather than a matching condition resulting in an enhanced phonological processing), means that further investigations are needed to clarify how beneficial rhythmic training might actually be for sentence processing. Since multimodal engagement with musical rhythm enhances rhythm perception (Geiser et al., 2010; Manning & Schutz, 2013; Su & Pöppel, 2012), it seems intuitive that such trainings would also impact on speech processing, though perhaps here a greater amount of audio-motor training would be required for a more apparent group difference to be seen. Similarly, a greater effect might also be observed if participants are made aware of the rhythmic relationship between the prime and the sentence, as they might be in a therapeutic context.

5.1 Perspectives

Whilst the current experiment has considered the effects of musical rhythmic priming on speech perception, we can also predict that speech production would benefit from rhythmic priming, and through two possible mechanisms. Auditory rhythm engages the motor system (Chen, Penhune & Zatorre, 2008; Grahn & Brett, 2007), and can provide a direct template for timely speech initiation and/or production (e.g. Cummins & Port, 1998). Second, it might be predicted that production may benefit from an enhancement in perception. Several lines of evidence indicate strong links between speech perception and production: speech perception engages areas involved in speech production (review, Galantucci, Fowler & Turvey, 2006; Geiser, Zaehle, Jäncke & Meyer, 2008; Liberman & Mattingly, 1985; Watkins, Strafella & Paus, 2003) which may reflect how the speech production system (motor) is recruited to understand and predict what might happen next (auditory) (Pickering & Garrod, 2007). In this case, an enhancement of perception will necessarily result in an enhanced production, and vice versa. The next step would therefore be to test the effect of rhythmic priming on speech production. This would have more relevance in the context of speech rehabilitation therapies, many of which already recognise musical rhythm as an important fluency-enhancing tool (e.g. Thaut, 2005).

Acknowledgements

This research was supported by a grant from EBRAMUS ITN (Europe BRAin and MUSic) (Grant Agreement number 238157). We thank Frederic Anguille, François Guimont, Bastien Intartaglia, Anaïs Llorens and Patrick Marquis for technical support and help with data collection.

References

Astésano, C. (2001). Rythme et accentuation en français: invariance et variabilité stylistique. L'Harmattan.
Barnes, R. & Jones., M. R. (2000). Expectancy, attention, and time. Cognitive Psychology, 41, 254–311.
Boersma, P., & Weenink, D. (1999). Praat: Doing phonetics by computer. Software. [Computer program]. Version 5.3.05, 〈http://www.praat.org/〉
Bolton; T. L. (1894). Rhythm. American Journal of Psychology, 1894, 6, 145-238.
Cason, N. & Schön, D. (2012). Rhythmic priming enhances the phonological processing of speech. Neuropsychologia.
Chen, J. L., Penhune, V. B., & Zatorre, R. J. (2008). Listening to musical rhythms recruits motor regions of the brain. Cerebral Cortex, 18(12), 2844-2854.
Cooper, G. & Meyer, L. B. (1960). The rhythmic structure of music. Chicago: University of Chicago Press
Cummins, F., & Port, R. (1998). Rhythmic constraints on stress timing in English. Journal of Phonetics, 26(2), 145-171.
Dilley, L.C. & McAuley, J. D. (2008). Distal prosodic context affects word segmentation and lexical processing. Journal of Memory and Language, 59, 294 – 311.
Ellis, R. J., & Jones, M. R. (2010). Rhythmic context modulates foreperiod effects. Attention, Perception, & Psychophysics, 72(8), 2274-2288.
Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. Psychonomic bulletin & review, 13(3), 361-377.
Geiser, E., Zaehle, T., Jäncke, L., & Meyer, M. (2008). The neural correlate of speech rhythm as evidenced by metrical speech processing. Journal of Cognitive Neuroscience, 20(3), 541-552.
Geiser, E., Sandmann, P., Jäncke, L., & Meyer, M. (2010). Refinement of metre perception–training increases hierarchical metre processing. European Journal of Neuroscience, 32(11), 1979-1985.
Gordon, R. L., Magne, C. L., & Large, E. W. (2011). EEG Correlates of Song Prosody: a new look at the relationship between linguistic and musical rhythm. Frontiers in psychology, 2.
Grahn, J. A., & Brett, M. (2007). Rhythm and beat perception in motor areas of the brain. Journal of Cognitive Neuroscience, 19(5), 893-906.
Guenther, F. H. (2006). Cortical interactions underlying the production of speech sounds. Journal of communication disorders, 39(5), 350-365.
Handel, S. (1989). Listening: An introduction to the perception of auditory events. pp. 383
Hannon, E. E., & Johnson, S. P. (2005). Infants use meter to categorize rhythms and melodies: Implications for musical structure learning. Cognitive psychology, 50(4), 354-377.
Hirst, D.J. & Di Cristo, A. (1998). A survey of intonation systems. In D. J. Hirst & A. Di Cristo (Eds.), Intonation systems : a survey of twenty languages (pp. 1-44). Cambridge: Cambridge University Press.
Iversen, J. R., Patel, A. D., & Ohgushi, K. (2008). Perception of rhythmic grouping depends on auditory experience. The Journal of the Acoustical Society of America, 124, 2263.
Jäncke, L. (2012). The relationship between music and language. Frontiers in Psychology, 3.
Jones, M. R. & Boltz, M. (1989). Dynamic attending and responses to time. Psychological Review, 96, 459–491.
Jones, M. R., Boltz, M., & Kidd, G. (1982). Controlled attending as a function of melodic and temporal context. Perception and Psychophysics, 32, 211-238.
Jones, M. R., Moynihan, H., MacKenzie, N. & Puente, J. (2002). Temporal aspects of stimulus-driven attending in dynamic arrays. Psychological Science, 13, 313–319.
Kusumoto, K., & Moreton, E. (1997). Native language determines the parsing of nonlinguistic rhythmic stimuli. The Journal of the Acoustical Society of America, 102, 3204.
Lappe, C., Herholz, S. C., Trainor, L. J., & Pantev, C. (2008). Cortical plasticity induced by short-term unimodal and multimodal musical training. The Journal of Neuroscience, 28(39), 9632-9639.
Large, E. W., & Jones, M. R. (1999). The dynamics of attending: how people track time-varying events. Psychological Review, 106, 119–159.
Large, E. W. & Palmer, C. (2002). Perceiving temporal regularity in music. Cognitive Science, 26, 1-37.

Levelt, W. J. (1999). Models of word production. Trends in cognitive sciences, 3(6), 223-232.

New, B., Pallier, C., Ferrand, L., & Matos, R. (2001). Une base de données lexicales du français contemporain sur internet: LEXIQUE™ //A lexical database for contemporary french: LEXIQUE™ . L'année Psychologique, 101(3), 447-462.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. Cognition, 21(1), 1-36.

London, J. (2012). Hearing in time. OUP USA.

Manning, F., & Schutz, M. (2013). "Moving to the beat" improves timing perception. Psychonomic Bulletin & Review, 1-7.

Pasdeloup, V. (2005). Figures et fond dans la scène prosodique: leur résistance face aux variations du débit de parole. In Symposium Interface Discours-Prosodie-IDP05, Aix-en-Provence, France.

Patel, A. (2008). Music, Language and the Brain. New York: Oxford University Press.

Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. Frontiers in psychology, 2.

Phillips-Silver, J., & Trainor, L. J. (2005). Feeling the beat: Movement influences infant rhythm perception. Science, 308(5727), 1430-1430.

Phillips-Silver, J., & Trainor, L. J. (2007). Hearing what the body feels: Auditory encoding of rhythmic movement. Cognition, 105(3), 533-546.

Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension?. Trends in Cognitive Sciences, 11(3), 105-110.

Pitt, M. A., & Samuel, A. G. (1990). The use of rhythm in attending to speech. Journal of Experimental Psychology: Human Perception and Performance, 16, 564–573.

Quené, H. & Port, R. F. (2005). Effects of timing regularity and metrical expectancy on spoken-word perception. Phonetica, 62, 1-13.

Roncaglia-Denissen, M. P., Schmidt-Kassow, M., & Kotz, S. A. (2013). Speech rhythm facilitates syntactic ambiguity resolution: ERP evidence. PloS one, 8(2), e56000.

Rothermich, Schmidt-Kassow & Kotz (2011). Rhythm's gonna get you: Regular meter facilitates semantic sentence processing. Neuropsychologia, 50, 232–244.

Schmidt-Kassow, M., & Kotz, S. A. (2009). Event-related brain potentials suggest a late interaction of meter and syntax in the P600. Journal of Cognitive Neuroscience, 21(9), 1693-1708.

Schultz, B. G., Stevens, C. J., Keller, P. E., & Tillmann, B. (2012). The implicit learning of metrical and nonmetrical temporal patterns. The Quarterly Journal of Experimental Psychology, (ahead-of-print), 1-21.

Snyder, J. S., & Large, E. W. (2005). Gamma-band activity reflects the metric structure of rhythmic tone sequences. Cognitive Brain Research, 24(1), 117-126.

Su, Y. H., & Pöppel, E. (2012). Body movement enhances the extraction of temporal structures in auditory sequences. Psychological research, 76(3), 373-382.

Thaut, M. H. (2005). Neurologic Music Therapy in Speech and Language Rehabilitation. In Thaut M. H. (Ed.), Rhythm, music, and the brain: Scientific foundations and clinical applications (pp. 165–178). New York: Routledge.

Trehub, S. E. (2001). Musical predispositions in infancy. Annals of the New York Academy of Sciences, 930, 1–16.

Watkins, K. E., Strafella, A. P., & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. Neuropsychologia, 41(8), 989-994.

Zatorre, R. J., Chen, J. L., & Penhune, V. B. (2007). When the brain plays music: auditory–motor interactions in music perception and production. Nature Reviews Neuroscience, 8(7), 547-558.

EXPERIMENT 3

Title: Rhythmic priming enhances speech production abilities: evidence from prelingually deaf children

Authors: Nia Cason [1,2], Céline Hidalgo [3], Florence Isoard [3], Stéphane Roman [4] and Daniele Schön [1,2]

(1) Aix-Marseille Université, INS, Marseille, France
(2) INSERM, U1106, Marseille, France
(3) La Timone Children's Hospital, CAMSP-DA, Marseille, France
(4) La Timone Children's Hospital, ENT Unit, Marseille, France

Corresponding Author: Daniele Schön
Corresponding Author's Institution: Aix-Marseille University
Corresponding Author's Email Address: daniele.schon@univ-amu.fr

Keywords: Music and Speech; Rhythm; Metre; Speech acquisition; Deafness

Abstract

After recent findings that rhythmic priming can enhance speech perception, we hypothesised that this would extend to speech production. We measured the influence of metrical priming on phonological production abilities in prelingually deaf children with hearing devices. Children had to repeat simple sentences that could be preceded or not by a rhythmical prime. Also the sentence could either match or mismatch the metrical structure of the prime. Results showed that Matching metrical prime conditions resulted in a greater phonological accuracy of spoken sentences compared to Baseline and Mismatching Conditions. This suggests that musical rhythmic priming can enhance phonological production via 1) an enhancement of phonological perception of the sentence target, and/or 2) an entrainment of speech motor rhythms through rhythmic listening alone. Overall, we interpret these findings as evidence of the ability for musical metre to engage domain-general expectations which can be utilised both in the perception and production of speech.

Introduction

Both music and speech are dependent on the temporal unfolding of auditory events. Of current interest is the temporal patterning of stressed and unstressed events, or ‘metre’. Though speech does not display the same degree of metrical regularity as does musical metre, an analogy can nonetheless be made: in both domains, there exists a patterning of prominent and less prominent elements. It is the ability for musical metrical patterns to parody those of speech, and the speech-enhancing consequences of this, which is of current interest.

Perceiving the metrical structures of speech are important for both language acquisition and speech perception. In infants, the perception of metrical structures in speech are said to appear at around 6 - 9 months (Jusczyk, Cutler & Redanz, 1993), the time at which infants begin to acquire their native language. During this important milestone of development, metrical structures provide important cues for speech segmentation which allows infants to identify individual words within the otherwise continuous speech stream (Morgan & Saffran, 1995; Johnson & Jusczyk, 2001). Once metrical ‘rules’ of one’s native language have been acquired, speech metre contributes to phonological (Pitt & Samuel, 1990), semantic (Magne et al., 2007; Rothermich et al., 2011) and syntactic (Schmidt-Kassow & Kotz, 2009; Roncaglia-Denissen, Schmidt-Kassow & Kotz, 2013) processing.

One view is that predictable speech metre results in an anticipatory allocation of attentional resources in time in order to optimise processing (Pitt & Samuel, 1990). This has been described in terms of a coupling of internal (neural) with external oscillators (stressed syllables, as well as syllable-sized units) in the speech domain (Arantes & Barbosa, 2010; Port, 2003) but more particularly in the music domain by the Dynamic Attending Theory, DAT (Jones & Boltz, 1989; Large & Jones, 1999). This theory predicts an enhanced processing of auditory events which occur at temporally-expected positions, supposedly through a more efficient processing at points at which attentional energy is higher (Jones, Moynihan, MacKenzie & Puente, 2002; review Peelle & Davis, 2012). Importantly, neural resonance is said to continue even after the rhythmic stimulus has ended. If predictability is indeed a driving force behind enhanced speech processing, then it might be expected that the predictability afforded by musical rhythm can further enhance this effect: musical rhythm affords more predictability than does speech rhythm.

Indeed, musical rhythm is thought to be an important driver of enhanced speech fluency in several speech therapies (e.g. Rhythmic Speech Cueing, Thaut, 2005; Melodic Intonation Therapy, Stahl et al., 2011, 2013; The Verbotonal Method, Guberina, 1963). Rhythmic priming studies have been used to empirically investigate the influence of musical rhythm on speech processing. For instance, Przybylski et al. (2013) found that presenting a regular rhythm can enhance the syntactic comprehension of subsequent sentences in specific language impaired and dyslexic children.). This suggests that short-term listening of temporal structures modulates attention and subsequently enhances linguistic processing, even after a delay. Rhythmic priming has also been found to enhance phonological processing of words: phoneme detection was found to be quicker for phonemes occurring at temporally expected positions. Moreover, electrophysiological measures showed that when speech metre did not match the musical prime metre, a mismatch negativity-like event-related potential (ERP) was elicited, which suggests that the word was heard in relation to, or as a continuation of, the rhythmical prime (Cason & Schön, 2012).

On the basis of these findings, we conducted a behavioural study in which we found that metrical priming can also enhance phonological processing of spoken sentences (Cason, Astésano & Schön, submitted). Heard sentences (containing a syllable-final phoneme target) were preceded by primes which either matched or mismatched the metrical structure of the sentence. We found evidence that Matching prime-sentence conditions result in an enhanced phonological processing of sentences, as measured by reaction times to a phoneme detection task. Further, we found that this effect was augmented for participants who received an additional period of rhythmic training. Considering the strong link between speech perception and production (review Galantucci et al., 2006), especially during speech acquisition (Callan, Kent & Guenther, 2000) we hypothesised that metrical priming can also enhance speech production. To this effect, we investigated the effect of metrical priming on the phonological sentence production in a population whose speech production ability relies on an interaction between speech perception and production: moderately, severely and profoundly hearing impaired children undergoing speech therapy.

Methods

Participants

16 deaf children took part in the experiment. All but two were prelingually deaf, i.e. deaf from birth or had experienced hearing loss before the age of 2 years. These children were undergoing speech therapy. Two children were excluded from final analyses due to an inability to produce the musical rhythm, resulting in a total of 14 children (11 female, average age 8.72 years, s.d. 2.19). They had received either unilateral cochlear implants (CIs) (n=3), bilateral CIs (n=3) Conventional Hearing Aids (CHAs) (n=5), both a CI and CHA (n=1) or Bone Anchored Hearing Aids (BAHAs) (n=2).

Degree of hearing loss ranged from moderate (n=5) to severe (n=3) to profound (n=6). This study was carried out in accordance with the provisions of the World Medical Association Declaration of Helsinki. Parents gave informed consent on behalf of their children, who completed a Baseline speech production task, a short training with the rhythmic prime and an Experimental Task. Children were compensated for their time with a gift.

Materials

Experimental trials consisted of a rhythmic prime sequence (to be reproduced on hearing) followed by a sentence (to be reproduced on hearing). Whilst the rhythmic prime had a xxXxxX structure, sentences following the prime could either have a xxXxxX structure (rhythmically matching prime-sentence pair), or a xxxXxX (rhythmically mismatching prime-sentence pair, see Figure 1). Both prime and sentences always had six sounds/syllables.

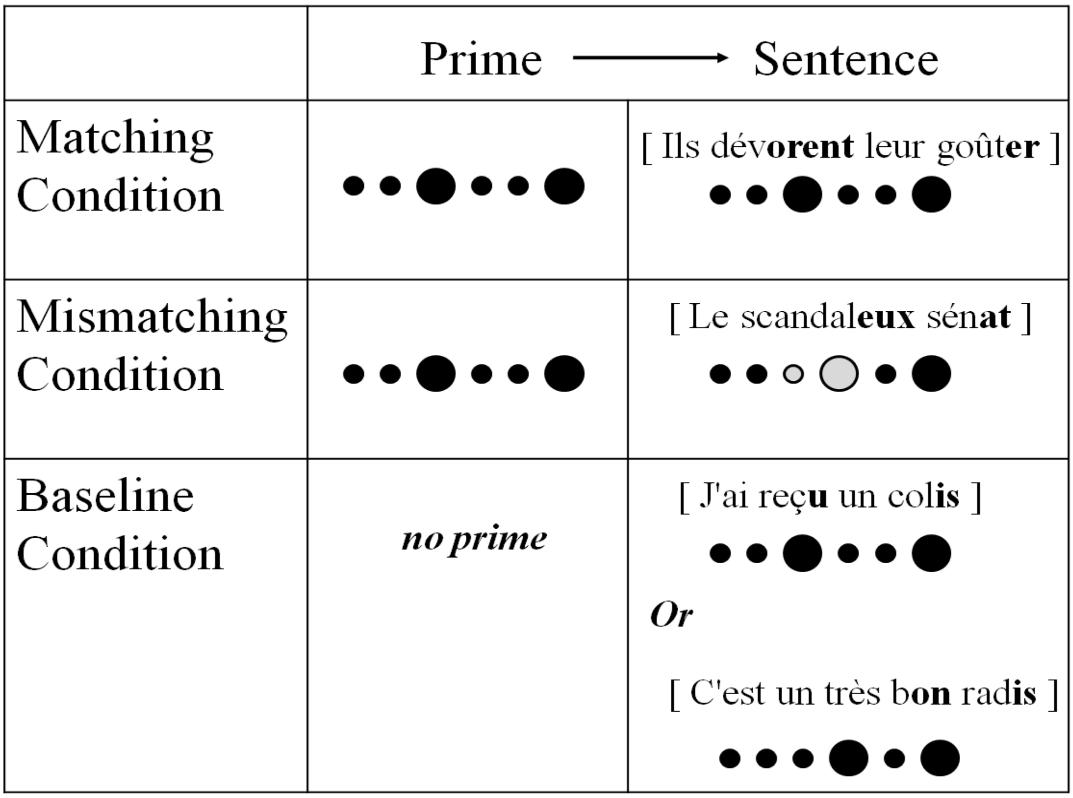


Figure 1. Schematic view of Matching, Mismatching and Baseline Conditions. Small and large circles indicate the weak/strong alternation of the metrical pattern in prime and following sentence. During Matching trials, a rhythmic prime (auditory rhythm to be reproduced) was followed by a spoken sentence with the same metrical structure as the prime. In Mismatching trials the sentence structure was not the same as the prime and the metrical difference is shown in grey. During Baseline trials there was no prime, and subjects had to repeat sentences with both of the above metrical structures.

The prime rhythm was the same used in a previous experiment (Cason, Astesano & Schön, submitted). It was created in Adobe Audition and was comprised of percussion sounds which had a stimulus onset asynchrony of 225 ms. The initial stressed percussion sound had a rim shot timbre, a duration of 196 ms and an average Root Mean Square (RMS) of -41.25 dB, the final stressed percussion sound had a snare timbre, a duration of 353 ms and an RMS of -28.88 dB, and the weak percussion sounds had a closed high-hat timbre, a duration of 138 ms and an RMS of -43.63 dB.

Sentence stimuli were also selected from a previous experiment (Cason, Astesano & Schön, submitted). These sentences either had the same metrical structure as the rhythmic prime (the position of stressed syllables mirrored the position of stressed beats of the prime, n=30), or had a mismatching metrical structure, whereby the syllable stress occurred one syllable later than would be expected according to the prime (n=30). Sentences varied in their number of vowels, consonants and words (mean no. of vowels = 6.01, s.d. 0.12; mean no. of consonants = 6.36, s.d. 1.007; mean no. of words = 4.4, s.d. 0.9). The number of vowels, consonants and words as well as word frequency was not significantly different for the two sets of sentence rhythms. These sentences were recorded in a sound-proof booth by a native French speaker and were spoken within a carrier sentence in order to control for sentence-ending prosodic effects. For instance, from the carrier sentence '"J'ai bu un bon cafe n'est pas facile a dire" [I drank a nice coffee is not easy to say], we extracted the first part of the sentence 'J'ai bu un bon cafe'. The last part of the carrier sentence had the same prosodic rhythm as the first (extracted) part of the sentence. Audio examples of the rhythmic prime, rhythmically matching sentences, and rhythmically mismatching sentences are available online as Supplemental Material.

Procedure

For the Baseline trials, children were asked to repeat heard sentences (of either a xxXxxX or xxxXxX structure) played by their speech therapist, focussing only on reproducing the speech sounds. Each sentence was heard only once and there were 20 trials. Sentences used in Baseline trials were pseudo-randomly assigned across subjects.

For Experimental trials, a rhythmic prime (xxXxxX metrical structure) was heard and reproduced using the sounds 'ti' and 'pa'. This was followed by a recorded sentence (with a xxXxxX structure (Matching) or xxxXxX (Mismatching) structure) which was also to be reproduced. In order to ensure the children were able to reproduce the rhythm, this was preceded by a short training with the rhythmic prime. Conditions order were randomised and counterbalanced across participants. The speech therapist presented the audio stimuli using the speech software

PRAAT, a professional USB sound card (Creative X-Fi Pro) and Hi-Fi Sony loudspeakers at a clearly audible level (adapted to each child) and recorded the whole session using an M-Audio Micro-Track II digital recorder. Due to the variability of speech abilities, individual performance determined the trial duration: the next trial was presented when the therapist deemed the child to be ready. Children performed the whole experiment with the same therapist.

Data Acquisition and Analyses

Speech production in deaf children is often measured by the percentage of accurately produced phonemes (Schriberg et al., 1997). We followed a similar strategy. For each spoken sentence, accuracy at the levels of the vowel, consonant, syllable and word were measured. These accuracy scores were calculated in terms of a % (proportion of accurately produced elements in relation to how many of these elements were present in the target sentence). Rhythm reproduction during experimental trials was also scored (as correct/incorrect metrical pattern), and only children who could accurately reproduce the rhythms were included in the final analyses.

Initial RM ANOVA tests were run on the data (separately for vowel, consonant, syllable and word levels). Post-hoc analyses reported are all Fishers tests. Multiple Regression analyses identified ‘degree of deafness’ to be a predictive factor of the main effect of condition. Further Friedman tests were therefore separately run on CI users (who comprised all children with profound deafness) and Hearing Aid (HA) users (i.e. CP and BAHA users). Further post hoc comparisons on these data were performed using Wilcoxon tests.

Results

Task Performance

All children were able to perform the sentence reproduction task well (mean percentage of missed trials: Baseline = 4.68 %; Experiment = 2.91 %). The mean % of incorrectly produced rhythms in the experimental trials was 2.09%. Missed trials (no response) received a score of zero and were included in the analyses. For each child, phonological accuracy was measured for Baseline and Experimental (Match and Mismatch) trials at the levels of the vowel, consonant, syllable and word. This score was converted into a % accuracy score, and the mean for each subject at each level and for each condition was calculated for the final analyses. For each of these levels (vowel, consonant, syllable and word accuracy), an ANOVA with 'Condition' as a within subjects factor with 3 levels (Match / Mismatch / Baseline) was run.

Greater Phonological accuracy in Matching Conditions

For phonological accuracy (%) there was a main effect of Condition, and at all levels (vowel: $F(2, 26)=4.99$, $p=.014$; consonant: $F(2, 26)=4.75$, $p=.017$; syllable: $F(2, 26)=4.37$, $p=.022$; word: $F(2, 26)=4.54$, $p=0.02$, (Figure 2).

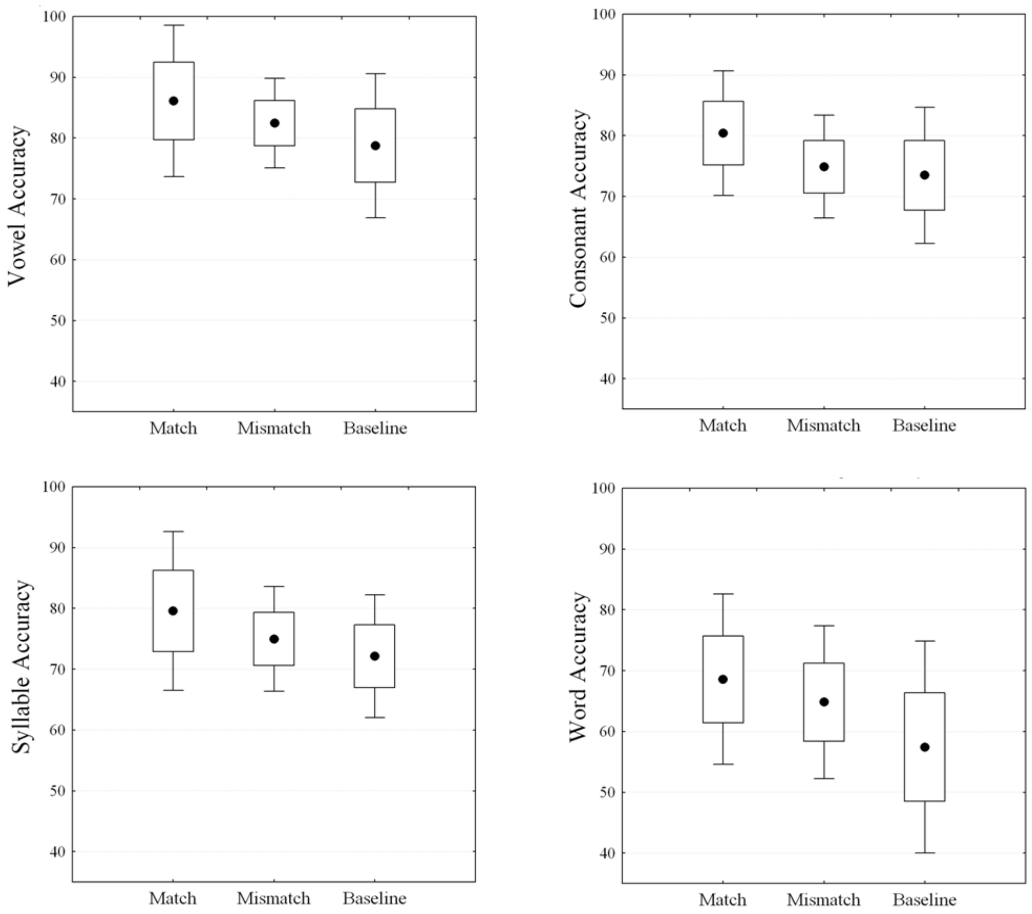


Figure 2: Phonological Accuracy (%) is shown for the Matching, Mismatching and Baseline conditions at all levels of analyses. The confidence intervals were computed as described in Loftus and Masson (1994) and take into account inter-subject variability.

Post-hoc analyses revealed that this effect arose from the difference in phonological accuracy in Matching vs. Baseline trials. This was significant at all levels (vowel p= 0.004; consonant p= 0.007 syllable p=0.0069; word p= 0.005). Matching Conditions also resulted in a greater phonological accuracy compared to Mismatching Conditions. Though this was not significant at the levels of the vowel (p=0.182) or the word (p=0.21), it was significant at the consonant level (p=0.028) and marginally significant at the level of the syllable (p=0.081). There was no significant difference between Baseline and Mismatching Conditions at any level (consonant p= 0.55; syllable p= 0.27), although the difference almost reached significance at the vowel and word levels (vowel p= 0.08; word p=0.09). In order to eliminate the possibility that xxXxxX sentences (Matching sentences in Experimental conditions) were simply easier than xxxXxX sentences (Mismatching sentences in Experimental conditions), we ran a paired t-test for production of these two sentence types within the Baseline. This revealed there to be no significant differences in phonological accuracy between the two sentence types, and at all levels (syllable p=0.47; vowel p=0.113; consonant p=0.419; word p=0.175).

Predictive Factors: Degree of Hearing Loss

Two multiple regression analyses were run in order to identify any predictive factors for the experimental effect for 1) Matching and Baseline Conditions (Matching - Baseline accuracy) and 2) Matching and Mismatching Conditions (Matching - Mismatching accuracy), with the factors 'age', ''type of device' (CI/CHA/BAHA/CI+CHA), 'degree of hearing loss' (moderate/severe/profound), 'duration had the device' and 'age started using device' as regression factors. This analysis was done on pooled data (vowel, consonant, syllable and word differences altogether). 'Degree of hearing loss' (Moderate/Severe/Profound) was predictive for the difference between Matching and Baseline Conditions (p=0.006) but there were no predictive factors surviving the threshold for the difference between Matching and Mismatching conditions. Overall children who had profound hearing loss showed a greater phonological accuracy in Matching vs. Baseline. This 'Profound' group was exclusively comprised of CI children. For this reason, another analysis was run which distinguished between CI users (n=7) and all other children (HA users, i.e. BAHA and CHA users) (n=7).

We ran two separate Friedman tests on CI users and HA users with the factor 'Condition' (Match/Mismatch/Baseline) as independent variable. CI children (female n=5, mean age 9.78 years, s.d. 2.7) had either profound (n=6) or severe (n=1) hearing loss, had received the implant at a mean age of 2.9 years (s.d. 1.82) and had been using the device for a mean of 6.8 years (s.d. 2.24). HA children (female n=6, mean age 7.5 years, s.d. 1.34) had either severe (n=2) or moderate (n=5) hearing loss, had received the implant at a mean age of 2.97 years (s.d. 2.89) and had been using the device for a mean of 4.59 years (s.d. 2.16). Rhythm reproduction errors occurred 5.71% of the time (s.d. 8.38) for the CI group and 2.14% of the time (s.d. 2.67) for the HA group.

For the CI group, there was a significant effect of Condition on phonological accuracy at all levels (vowel (Q (N = 7, dl = 2) = 8.8, p = 0.01); consonant Q (N = 7, dl = 2) = 7.14 p = 0.028; syllable: Q (N = 7, dl = 2) = 8.8 p = 0.01; word (Q (N = 7, dl = 2) = 10.28 p = 0.0058). For the group of HA users, there was no significant effect of Condition at any levels (syllable Q (N = 7, dl = 2) < 1; vowel Q (N = 7, dl = 2) = .074 p = 0.96; consonant Q (N = 7, dl = 2) < 1; word Q (N = 7, dl = 2) < 1) (Figure 3).

In order to see where these differences arose from within the CI group, we ran post-hoc comparisons (Wilcoxon test, M vs. MM / M vs BL) at all levels (vowel, consonant, syllable and word). This revealed a significant difference between Match and Baseline phonological accuracy at the level of the vowel (p=0.018), consonant (p=0.018), word (p=0.017) but not at the level of the syllable (p=0.23). Match and Mismatch Phonological Accuracy was significantly different at the syllable level (p=0.027) and word (p=0.027), almost reached significance at the level of the vowel (p=0.063) but not at the level of the consonant (p=0.13).

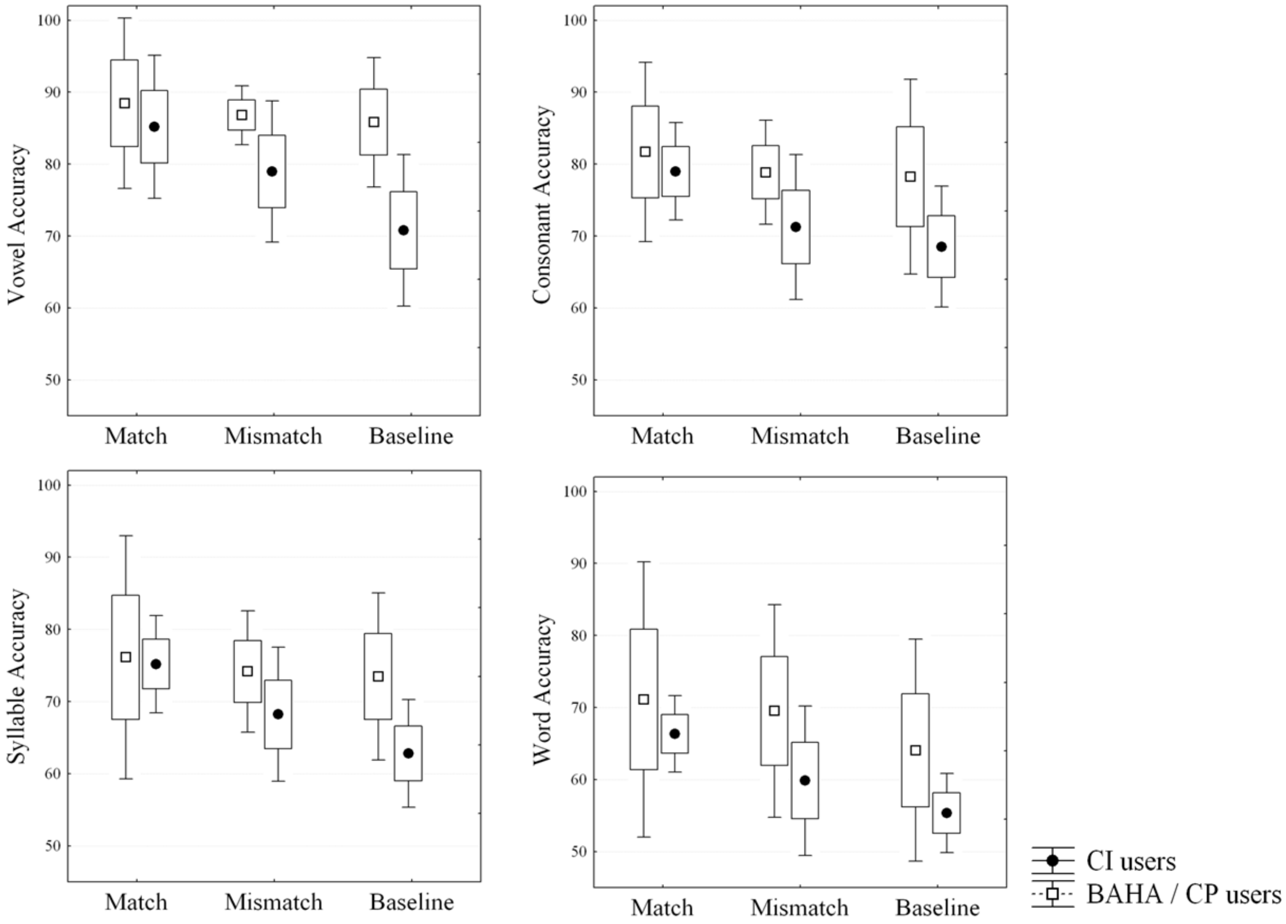


Figure 3: Phonological Accuracy (%) is shown separately for CI and HA users for the Matching, Mismatching and Baseline conditions at all levels of analyses. The confidence intervals were computed as described in Loftus and Masson (1994) and take into account inter-subject variability.

Discussion

1. Metrical Priming enhances phonological production

In this experiment, we investigated sentence reproduction performance when it followed a Matching rhythmic prime, a Mismatching rhythmic prime or no prime (Baseline). We found that priming well for speech metre enhances the phonological production accuracy of spoken sentences. In order to speculate by what mechanisms this effect came about, it is important to consider 1) the effects of predictability on speech perception, and 2) the link between perception and production of speech, specifically in relation to the speech abilities of prelingually deaf children. Based on these two lines of evidence, we propose that an enhanced speech production comes about through an enhanced speech perception of a heard target.

The online perception of speech benefits from rhythmic predictability. This predictability can be afforded by metrically regular speech (Pitt & Samuel, 1990; Rothermich et al., 2011; Schmidt-Kassow & Kotz, 2009), or by musical metre (Cason & Schön, 2012; Cason, Astésano & Schön, submitted). Due to its highly predictable nature, musical rhythm may have a greater capacity to enhance speech processing via this mechanism than speech. The DAT describes how a heard rhythmic structure can focus attention in time and consequently optimise processing resources, even once the entraining stimulus has ended (Jones & Boltz, 1989; Large & Jones, 1999). In addition, attentional allocation does not necessarily rely on isochrony (e.g. Pitt & Samuel, 1990): in this experiment, sentences were not aligned to the hypothetical beat inferred by the metrical rhythmic primes. We can therefore propose that metrical predictability resulted in an enhanced perception of target sentences, which consequently resulted in their enhanced production.

The interaction between speech perception and speech production during speech acquisition is an important one. The Motor Theory of speech (Liberman & Mattingly, 1985) claims that speech perception arises from the identification of the motor act involved in their production. Since then, much evidence has implicated the involvement of speech motor systems in speech perception, and vice versa (review Galantucci et al., 2006). Indeed, in speech acquisition the perception of self-made speech strengthens audio-motor associations required for correct phonemic articulation (production), and this in turn impacts on perception (Callan et al., 2000). Perception and production are also interdependent elements of spoken language in prelingually deaf children acquiring language (Blamey et al., 2001), distinguishing this population from other speech-impaired populations who experience an asymmetric or selective deficit of either perceptual abilities (e.g. post-lingual deafness or receptive aphasia) or of production (motor) abilities (such as stuttering or non-fluent-aphasia). Whilst the perception of syllable number is unimpaired (Boothroyd, 1984), hearing-impaired individuals have difficulties in perceiving stress patterns in speech (Jackson & Kelly-Ballweber, 1986; Most & Peled, 2007) This may also be tied to a general cognitive sequencing deficit; early sound input is thought to provide an auditory scaffold for cognitive sequencing across modalities (Conway et al., 2009). To note, the only difference between Matching and Mismatching conditions in this experiment was the similarity of prime and sentence metrical structures; though Matching sentences were consistently of a xxXxxX and Mismatching sentences of a xxxXxX structure, children showed no significant difference in phonological accuracy between the two sentence types in the Baseline condition.

Overall, we can hypothesise that a) considering the interactive nature of speech perception and production, metrical priming enhanced the ability to detect contrasts in speech, which in turn led to an enhanced phonological perception and thus production, and/or b) rhythmic reproduction during experimental trials (prior to sentence reproduction) familiarised children with the metrical structure of heard sentences and also strengthened entrainment of speech motor sequences required for the correct metrical production of the sentence. Based on findings that phonological perception of sentences is enhanced through rhythmic priming and that rhythmic training enhances this effect (Cason, Astésano & Schön, submitted), we propose that both these factors are valid explanations.

2. CI children best responded to rhythmic priming

The finding that only the speech production abilities of CI children benefited from rhythmic priming can be explained in one of two ways. This group was mainly comprised of children who were profoundly deaf, making it difficult to decipher whether this finding is due to the fact that a) this group of children had a greater hearing impairment, or b) this group had received CIs.

To begin with the first scenario, Baseline speech accuracy of CI users (profound deafness n= 6, severe deafness n=1) was lower than children who used HAs (severe deafness n=2, moderate deafness n=5). Since there were no group differences in the age at which the child received their hearing device, the age of the child, or the number of years since receiving the device, this lower speech performance may reflect the slower acquisition of language in CI/profoundly deaf children. Indeed, mean Baseline production for HA users was greater than CI children at all levels (16.37% greater at the vowel level, 10% at the consonant level, 9.73% at the syllable level, 7.86% at word level). This lower accuracy in the Baseline task may mean that CI users had more room for improvement to benefit from rhythmic cues.

The second scenario is more related to the use of the implant. Due to physiological issues of implanting within the cochlea, to technical issues with the implant and the interaction of the two, CI use results in a limited spectral perception. While HA users receive equal amplification of all auditory signals, CI users have a better perception of slow temporal patterns compared to fast spectral changes (Nie, Barco & Zeng, 2006). Thus, while HA users may have heard both the prime and speech at an equally salient level the rhythmic prime may have been more easily perceived than the speech stimuli in CI users. In these instances, CI users would have relied more on rhythmic cues than HA users might have.

CI users may therefore be a population who particularly benefit from rhythmic priming. Whilst other developmental speech disfluencies have been suggested to arise from general temporal processing deficits (Tallal, 1993; Overy, 2003), this is not the case in prelingually deaf children. Indeed, rhythm perception abilities in CI children appear to be normal; they are able to detect off-beat irregularities (Kim et al., 2010), duration irregularities, recognise metrical structures (Cooper et al., 2008) and detect tempo differences of rhythmical sequences (Kong et al., 2004) just as well as normal-hearing children. In addition, CI children rely more on rhythmic cues to perform pitch perception tasks (Kong et al., 2004). They may also rely more on rhythmic cues due to their impaired ability to perceive suprasegmental features of speech (such as intonation, syllable stress and word emphasis) compared to children with HAs (Most & Peled, 2007): rhythm provides an unambiguous cue for this. Indeed, rhythm perception in CI users is strongly correlated with speech perception (Fu, 2002).

To conclude, our results provide evidence that musical rhythm engages a domain-general temporal processing mechanism. An anticipation of the speech metre (via the prime) is thought to have resulted in a facilitated phonological perception, thus facilitating phonological production - an effect which may have been facilitated through rhythmic engagement with the prime. CI users may have benefited more from this due to their greater reliance on temporal cues, or possibly due to a larger room for improvement. The results presented here thus have not only strong clinical potential, but also theoretical implications about the shared neural processing of speech and music.

Contributions

D.S. and N.C. developed the study concept. All authors contributed to the study design. Testing and data collection were performed by C.H. N.C. and C.H. performed the data analysis under the supervision of D.S. D.S. and N.C drafted the paper, and C.H., S.R. and F.I. provided critical revisions. All authors approved the final version of the paper for submission.

Acknowledgements

This research was supported by a grant from EBRAMUS ITN (Europe BRAin and MUSic) (Grant Agreement number 238157).

References

1. Arantes, P., & Barbosa, P. A. (2010, May). Production–perception entrainment in speech rhythm. Proceedings of speech prosody, 1-4.
2. Blamey, P. J., Sarant, J. Z., Paatsch, L. E., Barry, J. G., Bow, C. P., Wales, R. J. & Tooher, R. (2001). Relationships among speech perception, production, language, hearing loss, and age in children with impaired hearing. *Journal of Speech, Language and Hearing Research*, *44*(2), 264.
3. Boothroyd, A. (1984). Auditory perception of speech contrasts by subjects with sensorineural hearing loss. Journal of Speech, Language and Hearing Research, 27(1), 134.
4. Callan, D. E., Kent, R. D., Guenther, F. H., & Vorperian, H. K. (2000). An auditory-feedback-based neural network model of speech production that is robust to developmental changes in the size and shape of the articulatory system. Journal of Speech, Language and Hearing Research, 43(3), 721.
5. Cason, Astésano & Schön (under review). Bridging music and speech rhythm: rhythmic priming and audio-motor training effect speech perception. Cognition
6. Cason, N. R., & Schön, D. (2012). Rhythmic priming enhances the phonological processing of speech. Neuropsychologia, 50, 2652-8.
7. Conway, C. M., Pisoni, D. B., & Kronenberger, W. G. (2009). The importance of sound for cognitive sequencing abilities: the auditory scaffolding hypothesis. Current Directions in Psychological Science, 18(5), 275-279.

8. Cooper, W. B., Tobey, E., & Loizou, P. C. (2008). Music perception by cochlear implant and normal hearing listeners as measured by the Montreal Battery for Evaluation of Amusia. Ear and hearing, 29(4), 618.
9. Fu, Q. J. (2002). Temporal processing and speech recognition in cochlear implant users. Neuroreport, 13(13), 1635-1639.
10. Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. Psychonomic bulletin & review, 13(3), 361-377.
11. Guberina, P. (1963). Verbotonal method and its application to the rehabilitation of the deaf. In *International congress on education of the deaf. Washington, DC: US government printing office* (pp. 279-293).
12. Jackson, P., & Kelly-Ballweber, D. (1986). The relationship between word and stress pattern recognition ability and hearing level in hearing-impaired young adults. The Volta Review, 88(6), 279-287.
13. Johnson, E. K., & Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. Journal of Memory and Language, 44(4), 548-567.
14. Jones, M. R., & Boltz, M. (1989). Dynamic attending and responses to time. Psychological review, 96(3), 459.
15. Jones, M. R., Moynihan, H., MacKenzie, N., & Puente, J. (2002). Temporal aspects of stimulus-driven attending in dynamic arrays. Psychological Science, 13(4), 313-319.
16. Jusczyk, P. W., Cutler, A., & Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of English words. Child development, 64(3), 675-687.
17. Kim, I., Yang, E., Donnelly, P. J., & Limb, C. J. (2010). Preservation of rhythmic clocking in cochlear implant users: a study of isochronous versus anisochronous beat detection. Trends in amplification, 14(3), 164-169.
18. Kong, Y. Y., Cruz, R., Jones, J. A., & Zeng, F. G. (2004). Music perception with temporal cues in acoustic and electric hearing. Ear and Hearing, 25(2), 173-185.
19. Large, E. W. (2000). On synchronizing movements to music. Human Movement Science, 19(4), 527-566.
20. Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. Psychological review, 106(1), 119.

21. Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. Cognition, 21(1), 1-36.

22. Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. Psychonomic Bulletin and Review, 1, 476–490.

23. Magne, C., Astésano, C., Aramaki, M., Ystad, S., Kronland-Martinet, R., & Besson, M. (2007). Influence of syllabic lengthening on semantic processing in spoken French: behavioral and electrophysiological evidence. Cerebral cortex, 17(11), 2659-2668.

24. Morgan, J. L., & Saffran, J. R. (1995). Emerging integration of sequential and suprasegmental information in preverbal speech segmentation. Child development, 66(4), 911-936.

25. Most, T., & Peled, M. (2007). Perception of suprasegmental features of speech by children with cochlear implants and children with hearing aids. Journal of deaf studies and deaf education, 12(3), 350-361.

26. Nie, K., Barco, A., & Zeng, F. G. (2006). Spectral and temporal cues in cochlear implant speech perception. Ear and hearing, 27(2), 208-217.

27. Overy, K. (2003). Dyslexia and music. Annals of the New York Academy of Sciences, 999(1), 497-505.

28. Peelle, J. E., & Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. Frontiers in psychology, 3.

29. Pitt, M. A., & Samuel, A. G. (1990). The use of rhythm in attending to speech. Journal of Experimental Psychology: Human Perception and Performance, 16(3), 564.

30. Port, R. F. (2003). Meter and speech. Journal of Phonetics, 31(3), 599-611.

31. Przybylski, L., Bedoin, N., Krifi-Papoz, S., Herbillon, V., Roch, D., Léculier, L., ... & Tillmann, B. (2013). Rhythmic auditory stimulation influences syntactic processing in children with developmental language disorders. *Neuropsychology*, *27*(1), 121.

32. Roncaglia-Denissen, M. P., Schmidt-Kassow, M., & Kotz, S. A. (2013). Speech rhythm facilitates syntactic ambiguity resolution: ERP evidence. PloS one, 8(2), e56000.

33. Rothermich, K., Schmidt-Kassow, M., & Kotz, S. A. (2012). Rhythm's gonna get you: Regular meter facilitates semantic sentence processing. Neuropsychologia, 50(2), 232-244.

34. Schmidt-Kassow, M., & Kotz, S. A. (2009). Event-related brain potentials suggest a late interaction of meter and syntax in the P600. Journal of Cognitive Neuroscience, 21(9), 1693-1708.

35. Shriberg, L. D., Austin, D., Lewis, B. A., McSweeny, J. L., & Wilson, D. L. (1997). The percentage of consonants correct (PCC) metric: extensions and reliability data. Journal of Speech, Language and Hearing Research, 40(4), 708.

36. Stahl, B., Henseler, I., Turner, R., Geyer, S., & Kotz, S. A. (2013). How to engage the right brain hemisphere in aphasics without even singing: evidence for two paths of speech recovery. Frontiers in human neuroscience, 7.

37. Stahl, B., Kotz, S. A., Henseler, I., Turner, R., & Geyer, S. (2011). Rhythm in disguise: why singing may not hold the key to recovery from aphasia. Brain, 134(10), 3083-3093.

38. Tallal, P., Miller, S., & Fitch, R. H. (1993). Neurobiological basis of speech: a case for the preeminence of temporal processing. Annals of the New York academy of sciences, 682(1), 27-47.

39. Thaut, M. (2005). Rhythm music and the brain: Scientific foundations and clinical applications (Vol. 7). Routledge.

Chapter 5

METHODOLOGICAL CONSIDERATIONS

The purpose of this (brief) chapter is to share some interesting methodological considerations which have been raised during this thesis. What will be outlined is here is far from being specific to the topic of this thesis, but relates to the broader issues in experimental research.

The interpretations of experimental results largely depends on methodological issues and the research methods employed. For instance, behavioural data (reaction time data, in the present case) is used to infer the difference between underlying cognitive processes, and the comparative speed with which these operations are carried out. EEG, on the other hand, can directly measure the cortical activity which may underly these processes, and at a high temporal resolution (often measuring the activity which precedes and may thus predict reaction time data).

Whilst data from behavioural and EEG methods may be complementary - in that behaviour may be causally linked to cortical activity - they differ in the fact that, whilst electrophysiology is able to capture on-line cortical responses as they occur (to the order of milliseconds), behavioural reaction time data rather captures the consequence of these operations and additionally involves an integration of these processes with a decisional process and a motor component (e.g. button press). EEG may thus also capture differences in cortical activity which are not necessarily observable in behaviour (e.g. François & Schön, 2010).

Focussing more on experimental design is to consider the basic groundwork of experimental research; experimental results are often interpreted in terms of a baseline or control condition. Strictly speaking, this means that findings from one experiment can not apply to any context outside that of the experimental one, and indeed, that interpretations may not be replicated should different (although just as 'valid') control conditions be adopted. For this reason, whilst experimental data can strongly imply an experimental finding (outside the context of the experiment itself), it must be interpreted with careful scientific vocabulary.

Experimental findings are also interpreted within a specific context. For example, studies may be conducted within a particular language (en especially relevant consideration within the realm of language research). Research may also take for granted theories, previous research, or the existing frameworks in which to interpret results: for instance, results which have been interpreted from a psycholinguistic perspective in which language is composed of phonology, lexical access, syntax and semantics, may equally be considered from a more neurocognitive perspective in which language could be considered in terms of its processing requirements (Besson & Schön, 2001).

Finally, experimental findings are dependent on the methods of analyses adopted. For instance, rhythm production data may be analysed in several different ways, all of which may tell us something different about sensorimotor synchronisation abilities (see Repp, 2005; Repp & Su, 2013). Some methodological issues concerning modes of analyses will now be considered in the context of the experiments carried out during this thesis.

A first consideration is the scoring of behavioural data. Whilst in Experiments 1 and 2 this measured reaction times to phonological detection (and as either 'correct' or 'incorrect'), Experiment 3 and the experiments of Appendices 1 and 2 measured i) speech reproduction abilities (Experiment 3, Appendix 2), and ii) rhythm reproduction abilities (Appendix 1).

In the first case (speech reproduction), phonological production accuracy in HI children is most commonly measured as a proportion of correct phonemes (Schriberg et al., 1997). In PWS, speech fluency is measured by standard tests (e.g. the Stuttering Severity Instrument 3) whereby sound repetitions, blocking (and, in some cases, accompanying facial tics) are assessed during the reading of two passages. Disfluencies for each syllable/word are then converted into a percentage score (Riley, 1994). Even in these two instances, there is a different approach; whist speech is measured in terms of correctly produced speech in HI children, it is commonly measured as incorrectly produced speech in PWS. Though one might expect these to reflect different sides to the same coin (i.e. that measuring incorrectly produced speech allows us to also ascertain the correctly produced speech), this is not the case. In the case of stuttering assessments, for instance, just because a word or a phoneme does not have a stutter-specific disfluency, that is not to say it has been correctly produced.

In Experiment 3, we measured phonological production accuracy at the vowel, consonant, syllable and word levels, and for each of the experimental conditions (sentences preceded by a matching, mismatching or absent prime). Each of these levels can be said to measure at different sensitivities articulation abilities of HI children. These scores were converted into a % score. A similar strategy was used in scoring speech production data in an experiment performed with PWS in which reproduction accuracy (of vowels, consonants, syllables and words) was scored (Appendix 2). The scoring of articulation accuracy (Experiment 3) and speech fluency (Appendix 2) is subjective. For this reason, it is advantageous for these assessments to be made independently by different scorers, in order to ensure that there is a scoring cohesion. Notably, however, it has been found that subjective scoring well-reflects computer-based automatised scoring, at least in terms of pitch performance (Larrouy-Maestri et al., 2012).

In the case of rhythm reproduction, we also measured this subjectively. In Experiment 3, HI children were additionally asked to reproduce the rhythmic prime (since audio-motor engagement with rhythm may enhance the priming effect, see Experiment 2). Prime reproductions were scored as either ‘correct’ or ‘incorrect’, and were assessed on the basis of whether the child correctly (vocally) reproduced the sequence of salient and less salient metrical events (using the sounds ‘ti’ and ‘ta’). Two children who were consistently unable to reproduce the rhythmic prime were excluded from the final analysis, on the assumption that this meant they could not perceive the metrical structure, which is considered to be a necessarily prerequisite for rhythmic priming (however, production may not necessarily predict perception (Sowiński & Dalla Bella, 2013)).

In another experiment (Appendix 1), native French and English participants were asked to reproduce French-derived and English-derived musical sequences, also vocally using the sounds ‘ti’ or ‘ta’. The hypothesis of this experiment was that native language rhythm impacts upon ability to reproduce musical rhythms (i.e. French and English-derived rhythms would be better reproduced/retained in memory according to a subject’s native language). Due to the durational complexity of the primes, the fact that we also measured reproduction on the basis of correctly reproduced ‘stressed’ and ‘unstressed’ events, and that tempos of reproduction differed from the original sequences, we also used a subjective method for scoring this data (a score from 0 to 9). Another option would have been to measure production (p-centre onsets) in relation to the precise metrical trace of the rhythms, and to measure the asynchronies from the original rhythm, for instance. However, as noted, this was not possible due to the high number of errors and variability (hesitations, tempo variability, errors in stress placement, errors in durational elements, missing or added elements).

In both these experiments, rhythm reproduction was measured via vocal rhythm reproduction. Whilst many studies in the field of sensorimotor synchronisation and reproduction abilities may rely on tapping data, the fact that participants vocally reproduced rhythms is not considered as a confound. Rather, it is a much more pertinent to a principle purpose of this research - speech. The underlying timing networks involved in rhythmic tapping production are probably much the same as those in rhythmic vocal production, the only difference being that it would recruit a different effector (the finger). Indeed, the motor control system underlying speech has been said to be the same for arm gestures (Gentilucci & Dalla Volta, 2008). In this case, we would expect similar results from both tapping and vocal rhythm production data (though a comparison of production abilities via different motor effectors could be interesting). Indeed, since the vocal apparatus is specialised for fine temporal motor production (i.e. speech) in all individuals, the greater accuracy with which rhythms are reproduced via this effector would maybe even better represent the efficiency of underlying timing networks. There may be cases, in which there is a speech-specific motor disorder (e.g. PWS). In this case, rhythmic ability may not only be better assessed through tapping data, but, from a clinical perspective, tapping may also engage networks involved in articulation, as has been proposed to be the case for non-fluent aphasic patients (Norton et al., 2009; Gentilucci & Dalla Volta, 2008).

Finally for methodological consideration is the choice of control conditions. In Experiment 1, a 2 x 2 design allowed us to measure the effects of on-beat vs. off-beat, and matching metre vs. mismatching metre. We found no interaction, but did find differences between on-beat vs. off-beat and mismatching vs. mismatching metre. To what extent can this tell us about the facilitating effect of predictability, as opposed to the interference effect of violated predictability? Nothing. In Experiment 2, we again measured the effects of a ‘matching’ condition to ‘mismatching’ conditions. This experiment also manipulated the type of prime-speech information which was mismatching (either stress patterns or the number of elements; metre being considered as a combination of the two) in an attempt to pinpoint which aspects of metre might be most salient in priming expectations. Broadly, reaction time data indicated that matching conditions resulted in an enhanced phonological processing than all mismatching conditions. Again, however, it may be the case that the difference between these conditions is due to an interference effect in mismatching conditions rather than a facilitatory effect in matching conditions. In order to speculate about the relative ‘facilitating’ or ‘interference’ effects, it may be important to have an additional control condition.

Although we do not report this within Experiment 1 and 2 articles, we did indeed include an additional control trial. The control ‘prime’ consisted of a single tone, or drum beat, who’s IOI (in relation to the onset of the word/sentence) was matched to those in experimental trials. This single-element control was thought to control for having a ‘warning’ signal for speech onset, so entrainment effects in experimental trials could be deduced. However, this particular control trial does not control for other aspects not directly related to entrainment to a musical rhythm, such as the number of elements or the time window within which auditory information is presented.

In the experiment conducted with PWS (Appendix 2), we designed a control condition which attempted to control for all factors except rhythmic entrainment. The control prime contained the same auditory information as both the rhythmic prime conditions (i.e. it had the same number of strong and weak metrical beats and the time window within which auditory events occurred was also the same), but in which the temporal order of the intervals between the control prime elements were disrupted so that it did not allow for the same degree of prediction. However, again, disrupting the prime’s temporal regularity/predictability could interfere with speech production. Conflicting representations (in this case, metrical representations of the prime vs. target sentence) may interfere with perception and production (e.g. as has been shown in phonological perception, Lukatela & Turvey, 1996).

In Experiment 3, matching and mismatching experimental conditions were also compared to a control condition. In this case, the baseline fluency of children was measured in the absence of a prime. Though this control condition does not control for other aspects, such as effects of general auditory input during the other two (with a prime) conditions, it can at least be said that there was no auditory information which interfered with phonological perception/production.

To summarise, combined EEG and behavioural research in may provide complementary data, and EEG may more sensitively detect cortical changes not necessarily apparent in behaviour. The use of subjective scoring has obvious downfalls but human error can be minimised by independent scoring. Importantly, the designs used in Experiments 1, 2 and Appendix 2 do not allow us to disentangle a facilitatory effect of temporal orienting from a inhibitory effect of unmet temporal expectations. Experiment 3 goes some way to determine the facilitatory effects of temporal orienting, comparing met/unmet expectations to absent temporal expectations. Interestingly, this seems to show that, even when a metrical expectations are not met, there is a facilitatory effect compared to absent temporal expectations. This could be explained by motivational factors, though.

Chapter 6
DISCUSSION

6.1 Overview of results

We found behavioural and electrophysiological evidence that temporal orienting can enhance the phonological perception of spoken pseudowords (Experiment 1) and sentences (Experiment 2). For behavioural data, participants in both experiments performed a phoneme detection task, with reaction times taken to represent the efficiency with which phonological information was processed. The goal of Experiment 1 was to investigate the effect of beat and metre expectancies on pseudoword processing. Using pseudowords means that 'purely' phonological processing is measured (i.e. there was no semantic/syntactic information present in the target speech). The goal of Experiment 2 was to investigate the effect of metrical priming on sentence processing, whereby a real sentence either matched or mismatched the metre of the prime. For this, Matching conditions consisted of prime-sentence pairs which were metrically identical in their position of stressed and unstressed elements, and (three) Mismatching conditions consisted of a prime-sentence 1) stress mismatch, 2) number mismatch and partial stress mismatch (only violated at the final two elements), or 3) a full stress and number mismatch. The purpose of these conditions was to more fully investigate the priming effect of two aspects of metre (number of elements and stress patterning). Since the temporal relationships between the prime and speech were different in these two experiments, they can be said to differ somewhat in the type of priming used. There may therefore be differences between the type of temporal processing systems engaged, which will be discussed below.

A further aim was to investigate how audio-motor (AM) training with rhythm can impact upon this cross-domain priming effect (Experiment 2). Not only did we find that AM training can augment the effects of rhythmic priming, but also that this was more prominent for primes which could more easily be reproduced during training. Taking into account this effect of AM training, in a subsequent priming experiment with HI children (Experiment 3), children were actively engaged with the rhythmic prime during experimental trials by not only listening to it, but also reproducing it.

In this experiment (Experiment 3), we found that metrical priming can enhance phonological production (i.e. articulation) in hearing impaired children with hearing aids (HAs) and/or cochlear implants (CIs). Interestingly, we also found that speech production of CI users was more greatly improved by rhythmic priming, compared to HA users. Overall, we hypothesised that an enhanced production occurred as a result of enhanced perception of the target sentence. This was due to the fact that, in this population, speech acquisition (and indeed, their impairment) is reliant on (or resulting from) the interaction between speech perception and speech production.

Broadly, these three experiments found that inducing rhythmic expectations can enhance phonological perception and production of speech. The following section (Section 6.2) will present the theoretical implications of these findings. It will consider these findings in terms of their place within existing theories of temporal processing, as well as their possible application to other linguistic levels of speech processing. Finally, it will consider one of the main hypotheses of this thesis: that similarities in temporal processes in speech and music may reflect domain-general cognitive processes. Clinical implications and future perspectives will be considered in Sections 6.3 and 6.4.

6.2 Theoretical Implications

6.2.1. How do findings relate to the current literature?

The finding that rhythmic priming can enhance phonological perception and production in a cross-domain manner suggests that music and speech share temporal processes. This section will discuss two possibilities in terms of these shared processes, and in light of the experiments conducted during this thesis.

The rhythmic primes used in these experiments (Experiment 1; Experiment 2; Experiment 3) differed in their complexity (Figure 6.1) and their temporal relationship with the subsequently presented speech. They may have, therefore, differed in their engagement of temporal processing systems for prediction.

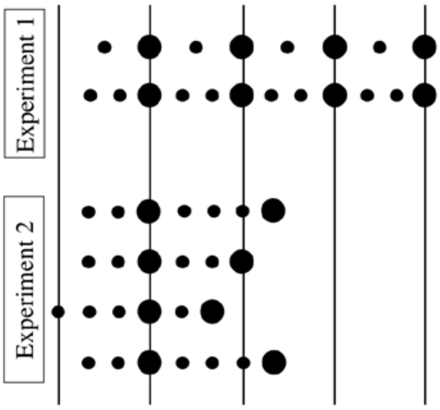


Figure 6.1
The rhythmic primes used in Experiments 1 and 2. Larger dots represent an acoustically salient event. Vertical lines show the regularity of salient events (the first salient event in Experiment 2 primes are aligned to the beat implied by Experiment 1 primes). For Experiment 1 primes, the beat and metre is regular (binary and ternary) and has four repetitions. For Experiment 2 primes, salient events do not occur at temporally predictable positions (with one exception). Interestingly, this exception (Prime 2 of Experiment 2) produced a different priming effect size.

In Experiment 1 we presented a binary or ternary metre (xXxXxXxX or xxXxxXxxXxxX) which was followed by a word which either metrically matched or mismatched this xX or xxX structure, and whose target phoneme was presented either ‘on-beat’ (at the expected 700 msec beat) or ‘off-beat’ (one syllable later).

In Experiment 2 and3, however, sentences following the prime did not conform to beat expectations; sentences were either 6 or 7 syllables in length with the stressed syllables occurring on either the 4th and final, or 3rd and final syllables. Because of this sentence structure variability, it was not possible to align them in a way in which the temporal interval between the prime and the sentence was controlled for (i.e. beat alignment was not possible). Instead, all sentences began at an equal distance from the prime and in such a way as to disrupt beat entrainment: 2.5 beats from the final stressed beat of the prime. Due to this, we measured metrical expectations only, independent of beat expectations.

Because of these differences in both the features of the primes used (greater length and repetition in Experiment 1) and their relationship to the speech (beat and metre / metre only), different temporal processes may be required for their encoding. In the first instance (Experiment 1), the metrical rhythms may induce an entrainment to both the beat and metre through a mechanism of neuronal resonance. The second instance might more involve the formation of abstract metrical representation. In this case, the second set of primes (Experiment 2) would draw more upon short-term memory; something which is not required by attentional entrainment.

To summarise, dynamic attending and/or abstract representation of temporal structure are two candidate temporal processes engaged by the rhythmic priming paradigms conducted during this thesis. These two possibilities will now be discussed in light of existing research.

6.2.1.1 Dynamic Attending

The online perception of temporal structures in speech and music have received very similar theoretical accounts; namely, that cortical oscillations attune to the periodicity of external auditory events, a mechanism which may be important for both prediction in time as well as for the grouping of auditory information. This account forms the basis of DAT in music (Jones & Boltz, 1989; Large & Jones, 1999; Large & Palmer, 2002), whereby different neurone ensembles (multiple oscillators) are said to synchronise with different metrical levels of a temporal pattern with phase and period relations which correspond to its hierarchical structure (see Figure 6.2). Through this mechanism, attentional resources are said to become focussed in time, to points at which an event is expected to occur - with more salient events thus receiving more attentional resources. This is said to optimise the brain's processing efficiency for auditory (and visual) information in a temporally dynamic way.

Whilst temporal structures in music are more regular than that of speech, this coupling of neural oscillators with external temporal structures has also been proposed for speech perception (Arantes & Barbosa, 2010; Giraud & Poeppel, 2012; Port, 2003) and production (Barbosa, 2002, 2007); multiple oscillators are said to attune to the periodicity of multiple levels of temporal organisation. These oscillators are thought to interact with one another in such a way that they are able to flexibly adapt their synchronisation to the temporal dynamics of speech, and are additionally thought to be sensitive to top-down predictive inputs - i.e. prediction which is based on linguistic knowledge (Zion Golumbic et al., 2012). In summary, attentional entrainment to both beat and metre is thought to occur through a mechanism of coupled oscillators, during both music and speech processing. However, due to the fact that beat and metre in music are more regular than in speech, temporal events in music may recruit oscillatory entrainment to a greater degree, thus allowing listeners to form stronger, or more automated, temporal predictions. The consequences of this will be discussed further in Section 6.2.4.

The finding that beat entrainment (through listening to musical rhythm) can enhance processing of speech events which occur at expected beat positions complements findings from both the speech and music literature. In the music domain, beat expectations have been found to enhance the processing of other musical dimensions, such as the processing of pitch intervals and melodic violations (Barnes & Jones, 2000; Jones et al., 1982, 2002; Bausenhart, Rolk & Ulrich, 2007; Jones, Boltz & Kidd, 1982), or implicit learning of grammatical pitch structures (Selchenkova, Jones & Tillman, in press). This work thus supports the view that attention is dynamically allocated in time, which results in a dynamic processing efficiency (DAT) (see Figure 6.2 for a hypothetical model of oscillatory activity in response to Experiment 1 primes). In the speech domain, The ABH similarly claims that phonological processing is enhanced at syllables which are expected to be stressed, due to a greater allocation of attention at these points in speech (Pitt & Samuel, 1990a). The authors found that priming listeners with a repetitive metrical context can positively impact on phonological processing when target phonemes occurred at positions which were predicted to be stressed, and they attribute this to a greater allocation of attention.

However, whilst Pitt & Samuel (1990a) presented metrically matching or mismatching words within a given metrical context (word lists), they did not present stressed syllables (which contained the target phoneme) at temporally isochronous positions - in which case, they may have been measuring the effect of metrical expectations, rather than, specifically, beat expectations.

A more recent study by Quené and Port (2005) aimed to further investigate the impact of beat and metre expectations on phonological processing. This study provided a predictable or unpredictable ‘beat’ contexts whereby stressed syllables (of bisyllabic words) were presented at isochronous intervals of 1.1 seconds, or not, with the target stressed syllable (containing the target phoneme) thus occurring at either expected or unexpected positions. The authors found that whilst beat expectations contributes to phonological processing, metrical expectations induced by a word list (iambic or trochaic word lists ending on a target word with the same/different metre) had no effect on phonological processing.

Similarly to their findings, whilst beat expectations did impact on phonological word processing, we did not find a behavioural (reaction time) difference between metrically matching and mismatching conditions in Experiment 1. However, we additionally measured EEG, which may be more sensitive that behavioural measures alone. We found electrophysiological correlates (a larger amplitude N100 for metrically mismatching vs matching conditions) which suggest that metrical expectations did indeed influence phonological processing. This larger early negativity in mismatching metre conditions could reflect the detection of a rhythmical deviance (Näätänen, 1990, also see Magne et al., 2007; Rothermich, et al., 2012) or, equally, a less efficient phonological processing. In addition, whilst Quené & Port (2005) used word lists to induce rhythmic expectations, we used a musical-like rhythmic prime, which may induce listener expectations to a greater degree than speech. Together, these findings also suggest that there is a distinction between beat and metre processing, which will be discussed further in Section 6.2.2.

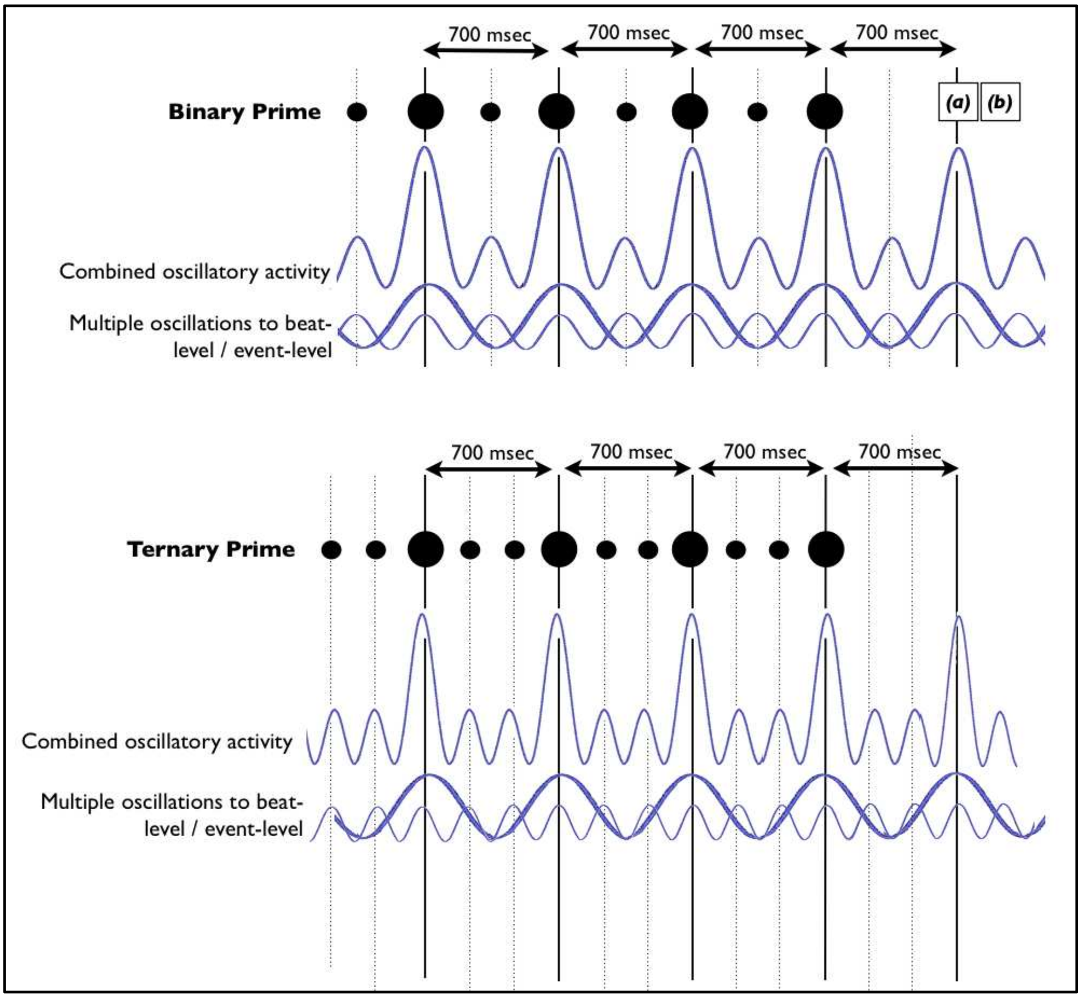


Figure 6.2
Hypothetical oscillatory activity in response to binary and ternary metres from Experiment 1 (based on DAT, Large & Jones, 1999). The summated activity of the coupled oscillators (resonating at either the beat/event levels) also represents the attentional allocation in time, and thus also processing efficiency. Auditory events (speech sounds in this experiment) occurring at windows of high attentional windows (e.g., at position (a)) are processed more efficiently than events falling outside this window (e.g. at position (b)). For simplification purposes, the gradual oscillatory synchronisation of period and phase is not depicted here.

6.2.1.2. Sequential Abstractions

In two subsequent experiments (Experiment 2; Experiment 3), we did not measure the effect of beat expectations. In these experiments, a prime was followed by a sentence whose sequence of stressed and unstressed syllables (metre) either matched or mismatched the metrical structure of the prime, but whose stressed syllables were not aligned to the hypothetical beat inferred by the prime. We thus measured the consequences of metrical expectation on sentence perception (via reaction time data in unimpaired participants, Experiment 2) and sentence production (via articulation accuracy of hearing impaired children, Experiment 3).

Our results indicate that, despite the absence of a ‘beat’ relationship between the prime and speech target, metrical expectations induced by musical rhythm can enhance phonological processing. This finding is in line with studies which have investigated the effect of speech metre (i.e. presenting metrically predictable or unpredictable word lists) on other aspects of linguistic processing such as syntax (e.g. Schmidt-Kassow & Kotz, 2009a; Roncaglia-Denissen et al., 2013) and semantics (e.g. Rothermich et al., 2012). It differs in two ways, however. Firstly, the metrical prime was presented in the form of musical metre (percussion sounds) rather than in the form of speech. Secondly, the prime we used provided a singularly-repeating metrical pattern, rather than a repetitive metrical context (see again Figure 6.1). Whilst both sets of primes are ‘metrical’, in that they imply a metric framework, the primes used in Experiment 2 may more more open to metrical interpretation, and only consist of 6 or 7 elements, pertaining to 2 or 3 beats - depending on the metric interpretation.

The question therefore arises as to whether cross-domain prediction in the metrical priming paradigm of Experiment 2 (and 3) can come about through neural resonance, since neural oscillatory activity and dynamic attending may require some degree of repetition. Indeed, there may be other temporal processes which come into play on hearing a more complex, or shorter metrical pattern to which one cannot so easily entrain to. One candidate for this could be the retention of the prime’s metrical structure in memory as an abstract memory representation. Indeed, the possibility that music perception and production requires different predictive strategies has been proposed by Keller (1999), who considers metre to provide both a framework on which to guide on-line temporal processing (temporal orienting) as well as to organise events in memory. He considers representational models of metre perception in which metre may be represented as symbolic codes (e.g. Povel & Essens, 1985), alongside models of dynamic attending (e.g. Large & Jones, 1999): considering that represenatationalist models reduce the role of metre to "that of a static template", he suggests that metre may have a 'dual' role, and that musical metre may be supported by both representational and procedural (online dynamic attending) processes. The idea that these representations are domain-general (and we can thus see an effect from music temporal structures to speech processing) will be discussed in Section 6.2.3.2.

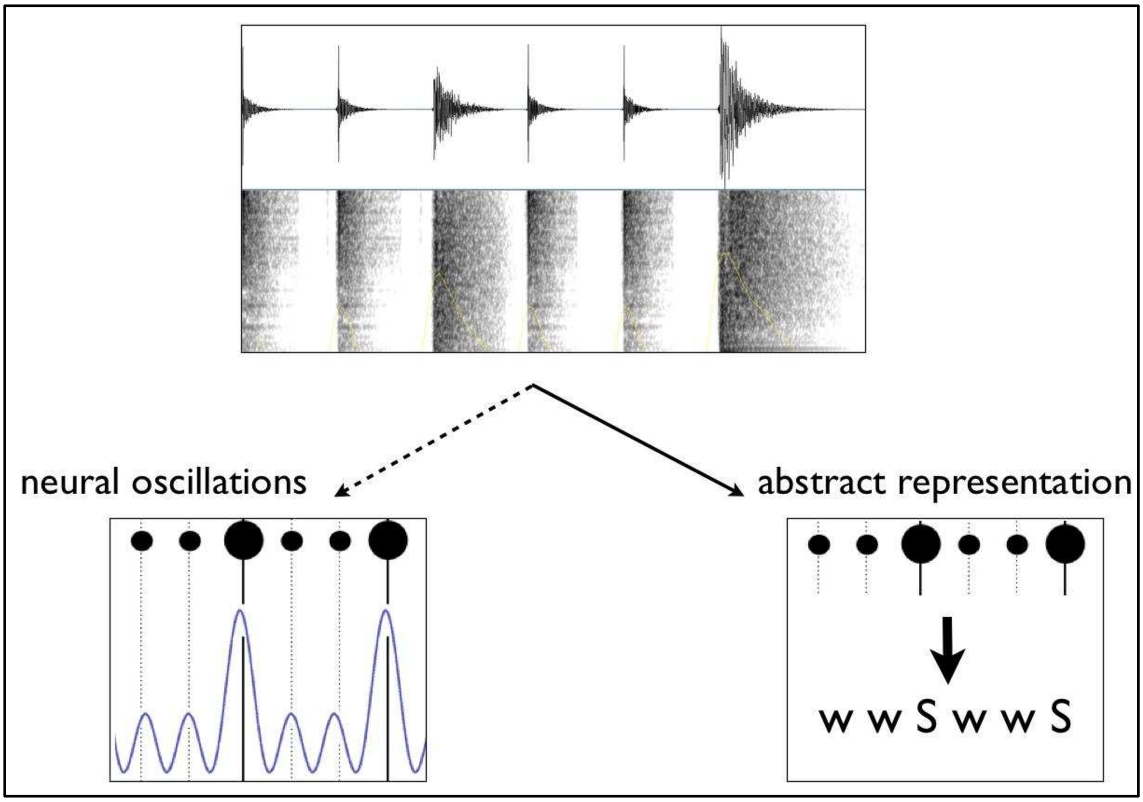


Figure 6.3
Two potential temporal processes engaged by a given temporal structure: attentional orienting and abstract representation. This example shows a prime used in Experiments 2 and 3 (with a xxX xxX metrical structure). Abstract metrical representations would also hypothetically be insensitive to tempo variations between the prime and speech, and may rely only on the metrical isomorphism of salient and less salient events (weak-weak-strong weak-weak-strong: w w S w w S). Differently metrical structures could engage these two processes preferentially according to their features. Here, this is indicated by the arrows; this particular metrical structure may preferentially rely on abstractions due to its non-repetitive structure (2 cycles only).

Overall, the metrical patterns used in the experiments presented here differ somewhat in both their predictability (i.e. a repetitive prime which could potentially induce synchronous neuronal oscillatory activity (Experiment 1) or a more metrically complex, shorter metrical prime which may rely on alternative temporal processes, such as storing the prime's metrical structure as an abstract, domain-general metrical representation (Experiment 2; Experiment 3) (Figure 6.3). Both these proposed temporal processes may be domain-general, in that we have observed in both instances a cross-domain effect of musical-induced prediction on speech processing. The extent to which metrical patterns of different complexities or lengths may engage in one or the other (or both) processes remains to be seen. It may also be the case that both processes are engaged during on hearing a temporal structure, but that one process benefits metrical structures to a greater degree, depending on the temporal characteristics/complexities of the stimulus. Alternatively, these two proposed processes may be two aspects of the same temporal process. These speculations would require further investigation.

This section has also outlined two further issues worth considering in more depth: firstly, that there may be a differential processing of beat and metre, or that beat expectations may be more salient than metrical expectations (Quené & Port, 2005; Experiment 1). Second, it may also be important to bear in mind that these observations are made within the realm of phonological processing; it is possible that different levels of speech processing may be more or less sensitive to different types of temporal expectancies. These issues will be addressed in the following sections.

6.2.2 Different mechanisms of prediction for beat and metre?

There is one clear distinction between beat and metre: whilst beat refers to one single temporal level, metre has, in addition to a beat, at least one additional overlaying structure. Whilst both beat and metre perception may recruit a mechanism of attentional orienting, metre perception may more greatly engage attentional orienting; it is supposed that multiple oscillators attuned to different metrical levels can result in a lower variability of oscillations responding to beat, thus reinforcing beat perception and expectations (Large & Jones, 1999; Jones, 2009). There may also be additional temporal processes involved in metre perception since metre perception requires hierarchical processing which is important for rhythmic grouping, whilst beat perception alone does not. These differences may suggest that metre perception recruits additional processes to that of beat which allow a listener to form not only expectations about ‘when’ events will occur in time (beat), but also ‘what’ events will occur in time (salient or less salient events).

The brain regions recruited during beat and metre perception in music may also give clues as to these differences. For instance, metrical processing is said to require greater short-term memory processes than hearing beat alone, as evidence by greater prefrontal cortex activation (a region strongly implicated in working memory processes; e.g. Courtney et al., 1997) during metrical vs. beat listening (Bengtsson et al., 2009), as well as in more complex metrical patterns vs. less complex metrical patterns (Sakai et al., 1999). Metre perception requires not only beat extraction, but also a hierarchical processing which may also facilitate the more cognitive feat of grouping of events into rhythmical units. Orienting attention in time through presentation of metre is said to recruit the IPC (Bolger et al., in press) and left IFG (Vuust et al., 2006). Temporal orienting through beat alone may additionally recruit the CB (Coull & Nobre, 1998; Coull, Frith, Büchel & Nobre, 2000). Whether or not there is a distinction between neural correlates of beat-based and metre-based temporal orienting has yet to be seen. This may be so, considering that beat and metre may induce different predictions (i.e. ‘when’ and ‘where’ contributions are different).

The idea that beat and metre may recruit different temporal processes can also be viewed from an angle of psychological timing. For instance, the different timescales which are required for different temporal processes may indicate that different temporal processes are involved in their perception. For instance, whilst an interval of 100 msecs is the lower limit required to hear precisely (with a 5-10% accuracy) durational differences of clicks (Hirsh, Monahan, Grant & Singh, 1990) and is the minimum interval required for for subjective rhythmisation to occur (Bolton, 1894; London, 2012), this timescale is not enough for beat perception, which occurs between 200 msec (just noticable difference) and 1500msec, with an optimal beat salience at 600 msecs (Fraise, 1982; Parncutt, 1994). However, since the IOIs required for subjective rhythmisation is 100ms, and for beat perception is 200ms, this has led to a hypothesis that hearing a beat requires at least the potential of hearing a beat division (London, 2012), in which case shared beat perception relies on the same temporal processes (in terms of millisecond timing) which are involved in metre perception.

Electrophysiology can shed light into the online processing of beat and metre as well as the effect of predictions induced by these temporal structures. Oscillatory beta-band activity (14 - 30 Hz) has been observed during beat and metre perception (Fujioka et al., 2009; Iversen, Repp & Patel, 2009), and neural resonance to metrical structures has also been marked by synchronised gamma band (> 30 Hz) activity (Snyder & Large 2005; Fujioka et al., 2009). In the current studies, however, we have measured ERPs in response to a conformation/violation of temporal expectations, since the particular interest has been the effect of music-induced expectations on speech processing. In Experiment 1, we measured the ERP responses to speech events which either conformed to the beat and/or metre of a rhythmic prime. Whilst beat conditions (on-beat/off-beat) impacted on P300 amplitude and latency, metre conditions (matching/mismatching) impacted on N100 amplitude.

The effect we found on N100 amplitude was interpreted as a violation of metrical expectancies. This has been previously found in the speech domain, with a larger early negativity recorded in response to metrically unpredictable words (Magne et al., 2007; Rothermich et al., 2012), as well as in the music domain, whereby violations of both pitch and temporal expectancies result in a larger N100 amplitude (Lange, 2009). The larger P300 amplitude was interpreted in light of classic interpretations of an increased amplitude being inversely proportional to the expectancy of a stimulus (Duncan-Johnson & Donchin, 1977; Squires, Wickens, Squires & Donchin, 1976; Tueting, Sutton & Zubin, 1970), and the earlier P300 latency observed in response to on-beat targets ties with evidence that targets occurring at attended timepoints result in a shorter P300 latency (Correa, Lupiáñez, Madrid & Tudela, 2006; Doherty et al., 2005; Griffin, Miniussi & Nobre, 2002; Schmidt-Kassow, Schubotz & Kotz, 2009). Due to the timecourse of these responses, we proposed that metrical expectations impacted on earlier, sensory levels of processing, whilst the effect of beat expectations seemed to impact on motor decisional processes. A similar interpretation is made by Lange (2009) who investigated conformations/violations of expectancy in the music domain; the smaller amplitude N100 found in response to pitch and temporal expectations which are met is similarly interpreted as an early modulation of expectancies on perceptual processing, and the similar effect of temporal expectations found on P300 amplitude was suggested to reflect the ability for temporal orienting to impact upon later cognitive (decision-related) or motor (response-related) processes.

The 2 x 2 factorial design (Experiment 1) allowed us to also investigate the interaction between beat and metre expectations. Whilst an interaction might be expected, due to the fact that metre is necessarily reliant on the presence beat, no interaction was found in either electrophysiological or reaction time data. However, it is important to remember that, rather than the processing of beat and metre themselves, we measured the effects of orienting attention in time (on-beat vs. off-beat) and sequencing (matched vs. mismatched metre). This absence of an interaction may be indicative of different neural bases of temporal orienting.

To summarise, temporal patterns with an increasing metrical complexity (i.e. those which have a clear beat structure and those which do not) may recruit similar brain regions, but to different degrees depending on the different processing requirements (e.g. a more complex metrical pattern may recruit memory processes to a greater extent than a simple metrical structure). However, of particular interest here the effect of temporal expectations on processing. For instance, whilst metrical expectations may impact upon early processes, beat expectations may impact upon later, decision- or motor-related processes. This may reflect the different strategies used. For instance, whilst beat alone (‘when’) may recruit regions involved in motor responses and synchronisation in time (e.g. Grahn & Brett, 2007), an overlaying metrical pattern (‘when’ and ‘what’) may be stored as a sensory memory trace, in which case a violation of metrical expectations impacts upon earlier, sensory processes. Metre may also recruit processes involved in sequencing and grouping information, and may rely more on short-term memory processes. In the case of phonological processing, our findings may also suggest that beat expectations are more salient than metrical expectations, as evidenced by the presence of both reaction time (to phoneme detection) and EEG data differences in beat conditions (whilst only EEG data differences were found in metrical conditions). As noted in Section 2.1, different complexities metrical patterns may also recruit different temporal processes, potentially related to memory components which are recruited to a greater extent when a metrical pattern offers less predictability. Overall, however, studies investigating the predictive effects of beat vs metre are lacking. Further investigations would be required to see to what extent metrical and beat expectations are interactive, or what processes they may recruit in order for prediction to come about.

6.2.3 Domain- and modality-general processes of cognition?

One interpretation of the work conducted during this thesis is that music may engage temporal processes which are also recruited during speech perception. This would explain how temporal expectations induced by musical rhythm can influence upon the perception of subsequently presented speech which either conforms or does not conform to temporal predictions. Whilst we have found that temporal prediction benefits phonological processing (be it via a mechanism of dynamic temporal attention or abstract representation), it is possible that different levels of speech processing (such as phonology, lexicon, syntax and semantics, which are defined by their characteristics) may be more or less sensitive to different temporal processes engaged by music.

However, another way to approach these findings is to look at how temporal prediction (rhythmic priming) can impact on the processes involved in these different levels of speech perception (e.g. sound analysis, retrieval from memory or rule extraction, for example): evidently, these processes are also necessary for music perception. For instance, both music and speech are said to require a processing of sound, structure and meaning (Slevc, 2012) - these processes are not only thought to be domain-general, but probably also interact with one another. In this case, our findings that phonological processing can be enhanced through temporal orienting may be comparable to findings that pitch processing can be enhanced through temporal orienting (Jones et al., 2002) - in that the underlying processes required for both phonological and pitch processing could be the same. Importantly, this view allows a comparison between speech and music to be made in terms of sensory and cognitive processes (Besson & Schön, 2001).

Furthermore, it may be the case that attentional orienting and the formation of abstract representations are not only domain-general feats of cognition (i.e. involved in both speech and music processing), but also modality-general. Whilst it seems clear that the auditory modality is specialised for the processing of temporal information, findings that attentional orienting is a cross-modal phenomenon (e.g. Bolger et al., 2013), and that the auditory rhythmic input is important for the development of sequential cognition in general (Conway et al., 2009) may support this view. These claims will be expanded on below. From this broader perspective, the focus becomes not ‘how music and speech may share resources’, but ‘how music and speech similarly draw upon cognitive processes’. In sum, by unifying evidence whose primary concern has been within a specific domain or modality, it may be possible to speculate about the engagement of domain-general general cognitive processes - two of which may be attentional orienting and abstract memory representations.

6.2.3.1 Temporal orienting

The presence of similar attentional accounts for the online prediction involved in both speech and music processing may reflect domain-general principles of cognition. In this case, speech and music may utilise these cognitive processes, but perhaps to a different extent (due to different processing demands). More explicitly is to say that music is more temporally regular than speech. This could suggest that predictability in music is more likely to come about through a process of attentional orienting, and, whilst speech may also rely somewhat on attentional orienting processes, prediction in this domain may rely more on implicit knowledge about a language's metrical patterns.

Many studies investigating the metrical processing in speech tend to use speech constructs which are metrically regular, but which one would not normally expect to occur in everyday conversational speech. Whilst this may be important to address the experimental question at hand, this may also mean that there is a recruition of attentional processes which may not normally be engaged in spontaneous speech with a less predictable metrical structure. For instance, a metrically regular speech construct (e.g. with a repeating trochaic metrical structure) would be expected to induce temporal attending to a greater degree than a more 'natural' speech construct, whose prediction might rely more on knowledge-based, top-down inputs.

Temporal prediction is not only required during music and speech perception. It is a necessary requirement for all structured human behaviour, particularly also in movement. Notably, auditory temporal structures induce synchronised movement, allegedly through synchronised neural oscillations. Whilst this type of spontaneous motor synchronisation is most apparent in the music domain, this may be due to the regularity afforded by music, rather than through a music-specific mechanism. In 2 experiments, Dalla Bella et al. (2013) approached the question as to whether motor synchronisation was specific to music, or whether motor synchronisation arises through entrainment to regularity, no matter what the domain (as noted by Lidji et al., 2011a, 2011b). They investigated to what extent distractors (either temporally regular music or speech) impacted upon synchronised tapping to a metronome (with the idea that a greater distraction would disrupt tapping synchronisation to a greater extent) (Experiment 1) and also to what extent pitch structures may play a role in synchronisation (with the distractos of sung 'la's', sung lyrics, or spoken lyrics, all with the same degree of temporal regularity) (Experiment 2). Interestingly, they found that music did not provide a greater distractor than speech. Moreover, metronome synchronisation during pitch distractors was not greater compared to lyrics only distractors. The ability for music to induce motor synchronisation is therefore considered to be due to it's regular nature, to which synchronised neural oscillations can preferentially entrain to. In addition, this may not necessarily rely on musical melody. Overall, these findings strongly suggest that similar temporal processes (which allow for motor synchronisation), such as neural entrainment, are recruited by both domains.

In sum, there is evidence to suggest that both speech and music can orient attention in time, and thus induce motor synchronisation. However, since natural speech contains less temporal regularity than musical structures, the degree to which the two domains engage these processes may differ.

Though the auditory modality seems specialised for processing temporal information, that is not to say that temporal orienting is specific to the auditory system. Evidence that temporal orienting is a general feature of cognition which may span across not only domains, but also modalities, comes from research showing that visual rhythms can induce temporal orienting (albeit to a lower temporal precision than auditory stimuli) (Coull & Nobre, 1998; Grahn, 2012b). Additionally, temporal orienting induced via an auditory prime can cross-modally impact on the processing of visual (Bolger, Trost & Schön, 2013; Brochard et al., 2013; Doherty et al., 2005; Correa & Nobre, 2008; Escoffier et al., 2010; Miller et al.,, 2013) and tactile (van Ede, Lange, Jensen & Maris, 2011) information. However, intra-modal priming (auditory-auditory) seems to show an stronger advantage for temporal orienting (Bolger et al., in press). Finally, attentional orienting via both visual (Coull & Nobre, 2008), visuo-spatial (Coull & Nobre, 1998) and auditory information (Bolger et al., in press) has been associated with IPC activity. This is suggestive of a general system in attentional orienting (in both time and space).

6.2.3.2 Sequential cognition and abstract representation

Another temporal process potentially engaged by metre is the formation of abstract representations, which may be considered as one important aspect of sequential cognition. Parallels between sequential cognition in music and speech is considered in a recent paper of Tillman (2012). In this paper, she highlights evidence which suggests that both music and speech may recruit general cognitive processes for the sequencing of both tonal and temporal information. She also considers temporal orienting of attention as a process which can influence cognitive sequencing. Indeed, whilst dynamic attending and sequential cognition are considered separately here, there is no reason to suggest that temporal orienting is separate from sequential cognition, or even that temporal orienting itself is not a mechanism of sequential cognition. However, sequential cognition is a rather broad term, referring to the cognitive structuring of sequences in time.

Dominey et al., 2003 distinguish between three forms of sequential structure; serial structure (the relationship between elements), durational structure, and abstract structure. Evidently, these sequential structures are important in both music and speech. Of particular interest here is abstract structure. To present an example from Dominey et al. (2003), the two sequences ABCBAC and DEFEDF have different serial structures but the same abstract structure (123213) and are thus called isomorphic. In music, this type of abstract isomorphism can be seen in, for example, musical themes which are repeated in a different key. Indeed, it has been found that listeners are able to retain pitch interval and rhythm information (abstract, long-term memories) and recognise transposed melodies (Dowling & Bartlett, 1981). An example of this abstract isomorphism in speech might be a sentence which is spoken by two different people (who have different voice timbres/pitch). In both instances, a listener is able to recognise that the repeated theme/sentence has already been heard, but in another tonal context: listeners have retained an abstract representation of the original in memory, which is immune to a tonal shift. This abstract encoding also applies to temporal aspects, whereby a listener has no trouble to recognise music and speech as containing the same information when it occurs at different tempos. In this case, it is the relative relationships between elements which is significant to the listener, rather than the absolute values (Hulse & Page, 1988). This being said, absolute memory representation for tempo/rate in both speech and music probably also exist (Jungers et al., 2002).

In a priming scenario, a prime structure with a xxXxxxX structure followed by a matching speech metre of a xxXxxxX pattern would present listeners with a isomorphic abstract structure. As noted, and in relation to the research performed during this thesis, a musical metrical structure such as those presented to listeners in Experiment 2 and 3 might be have been retained in its abstract form. Indeed, temporal orienting during these sequences is theoretically less than during the presentation of a more repetitive metrical structure such as those presented to listeners in Experiment 1. In this case, a sentence following the prime was heard as either conforming to the abstract temporal structure or not. As well as across-domains, the formation and utilisation of abstract metrical representations may also be a general cognitive process which spans modalities. Indeed, Dominey et al. (2003) review evidence that agrammatic aphasic patients (who have a syntactic processing deficit) are also impaired in other non-linguistic abilities which require sequential cognition, and thus consider sequential cognition as something which encompasses not only perceptual but also motor sequencing.

Support for the view that temporally structured information can be represented in an abstract form, and thus relayed to other modalities comes from Conway et al. (2009). The authors propose that our capacity for hierarchical/sequential processing - across modalities - is provided by auditory input during development. In this ‘Auditory Scaffold hypothesis’, the authors propose that cognitive sequencing abilities may develop via early rhythmic auditory input. This hypothesis is based on evidence from individuals who have experienced early hearing loss (and thus who have not received this early auditory input). For instance, deaf children have other, non-auditory disturbances in temporal sequencing functions (Marschark, 2006); these children are impaired in motor sequencing tasks (Conway et al., 2011), for instance. However, it may be the case that other inputs (such as rhythmic tactile inputs) may compensate for loss in the auditory modality (Brochard et al., 2008), and that the development of sequential and hierarchical cognition can potentially occur via other modalities. On the other hand, visual and tactile cues are not considered to provide the same degree of encoding and processing accuracy in hierarchical and temporal relations as is the auditory system (Conway & Christiansen, 2005).

Overall, forming abstract representations of hierarchical/sequential (metrical) structures may allow not only a cross-domain isomorphism to occur (e.g. between music and speech metre), but perhaps also a cross-modal isomorphism. The auditory modality may provide the most efficient input for temporal information due to its greater temporal precision, and, via abstract representations which can span modalities, may provide vital inputs important for the development of cognitive sequencing abilities in general.

To conclude, musical metre perception may engage both dynamic attending through an online temporal orienting, and may additionally support the formation of abstract representations (e.g. Keller, 1999). Under the assumption that these processes are also subserved by temporal structures in speech (i.e. domain-general), rhythmic priming through music is hypothesised to engage these domain-general processes in such a way that speech processing 1) can be processed more efficiently according to it's temporal position relative to dynamically-allocated attention (dynamic attending), and 2) can be made predictable through a metrical isomorphism, or template, provided by music, which - due to its representation in an abstract form - can be applied to speech. Not only this, but these two proposed processes may represent general cognitive processes which are also modality-general. The utilisation of these two predictive processes may occur in parallel, or may even be complementary, and are probably recruited to different degrees depending on the processing requirements of the temporal stimulus.

6.2.4 Contribution of temporal prediction at different levels of speech processing?

In the experiments described, we have measured the impact of temporal expectations on phonological processing, which forms the building blocks of speech perception. From a neurolinguistic perspective, speech can be decomposed into constituent parts which, together, characterise language: prosody, phonology, lexicon, syntax, semantics and pragmatics. Timing plays an important role at each of these levels, which are additionally interacting in both a bottom-up and top-down way. Although another view - already considered - is that music and speech rely on more general cognitive processes (and that these processes could be more or less important for each of these assigned linguistic levels), this section will refer to the different linguistic levels of processing as they have been widely studied.

There is a potential for temporal processes to impact upon other processes other than phonological processing. The idea that different temporal structures (and therefore, potentially different temporal processes) may contribute differently to these different levels of speech processing will now be considered.

Metrical or beat regularity of speech seems to enhance the processing of phonology (Pitt & Samuel, 1990; Quene & Port, 2005), lexical access (Brochard et al., 2013; Grosjean & Gee, 1987; Dilley & McAuley, 2008) syntax (Roncaglia-Denissen et al. 2013; Schmidt-Kassow & Kotz, 2009a) and semantics (Astésano et al., 2004; Magne et al., 2007; Rothermich et al., 2012). However, in the same way music and speech may recruit different temporal processes to different extents depending on their temporal characteristics, it is also possible that different levels of linguistic processing are more or less sensitive to different sources of temporal prediction. For instance, whilst phonological processing (Pitt & Samuel, 1990; Experiment 1) or lexical access (Brochard et al., 2013) may be enhanced by inducing beat expectations, it may be the case that syntactic and semantic processing is more sensitive to, for instance, metrical priming (e.g. Schmidt-Kassow & Kotz, 2009a; Rothermich et al., 2012).

First for consideration is the secondary effects of rhythmic priming which may come about through the interactions between different levels of speech processing. Whilst is known that different levels of speech processing interact with one another, the precise relationship between these levels is not yet fully understood. For instance, it could be the case that whilst one, specific level of speech processing may directly benefit from rhythmic priming, this may also have a secondary impact upon the processing of other levels of speech. For example, an enhancement of phonological processing via beat induction may, as a consequence, impact upon other levels of processing, since phonological perception is required for lexical access, lexical access for syntactic processing and so on. Indeed, an enhanced phonological processing is a possible alternative explanation for findings that lexical access is enhanced through beat orienting (Brochard et al., 2013) or through an alignment of syllables with musical rhythm (Gordon et al., 2011). Alternatively, an enhancement of phonological processing via beat induction may - hypothetically - occur via an enhancement in syntactic processing, since a top-down effect in speech perception also occurs. The fact that metrical expectations impacted on early phonological processing (ERP evidence), so quite early on in the auditory cortex, also suggests that lexical processing or retrieval could benefit. This might be a future research question.

This uncertainty with regards to the interactions of different levels of speech processing may also explain why there are mixed views about the underlying deficits of language-impaired individuals. For instance, whilst SLI and dyslexic children are known to have a syntactic processing deficit, it is possible that this may arise as a consequence of impaired phonological processing (Georgiewa et al., 2002; Snowling, 1981). Overall, these linguistic interactions would have to be better characterised in order to determine from which level there may be a root benefit or deficit.

Next for consideration is the possibility that multiple levels of speech processing may benefit from rhythmic priming, and that different linguistic levels are more or less sensitive to different temporal structures/processes. Indeed, we have found a differential effect of beat and metre expectations on phonological processing (Experiment 1). Whilst both beat and metre expectations elicited different ERP components for the different (Matching/Mismatching) conditions, only beat expectations impacted upon phoneme detection (reaction time) data. This showed that whilst the word metre was indeed processed in relation to the prime metre, this does not necessarily have an impact on the detection of a target phoneme. This well reflects behavioural findings by Quené & Port (2005) who, similarly, found that beat expectations (induced through word lists) impacted on phoneme detection, but metrical expectations (iambic or trochaic words) did not. Metrical expectations may have a more robust impact on syntactic processing, for instance, where behavioural evidence is apparent, as well as electrophysiological evidence (Roncaglia-Denissen et al., 2013; Schmidt-Kassow & Kotz, 2009a). Syntax requires the grammatical organisation of words within a given context; this type of hierarchical processing may be more analogous to the hierarchical processing induced by metrical sequences. Indeed, it has been proposed that the syntax of an utterance (regardless of the words, semantics or context) can generate an underlying “baseline” prosody

(Arnfield, 1994). In this case, processes engaged by a musical metrical prime may engage similar processes recruited during syntactic processing, thus priming a listener for syntactic constructs more than beat might.

The SSIRH considers musical and linguistic syntax to rely on shared resources for structural integration. However, it may be the case that even more general cognitive processes, such as integrating information into an existing context, may be responsible (Slevc et al., 2009; Tillman, 2005, 2012). This may be supported by findings that musical syntax can influence linguistic semantic processing (Poulin-Charronnat et al., 2005; Steinbeis & Koelsch, 2008) - semantic processing also requires information integrations, whilst may be less reliant on structural integration. Alternatively, this may be explained by the interactions between syntactic and semantic linguistic processing (Friederici, 2002; Friederici & Kotz, 2003), whereby an impact on syntactic processing can impact upon semantic processing.

Electrophysiological markers of could also give us a clue as to what levels of linguistic processing are effected by temporal prediction. For instance, in Experiment 1, we observed an N100 effect of metrical expectations which were or were not met by speech. Though we interpret these findings as a detection of rhythmical incongruity, it is worth noting that the N100 is a marker for phonological processes in speech (e.g. review Friederici, 2002; Friederici & Kotz, 2003). In this case, the larger N100 observed for matching (vs. mismatching) metrical conditions could reflect the greater resources required for phonological processing under these conditions. It could also be interpreted as an MMN-like response elicited in response to deviants in an auditory sequence; the N100 and MMN responses have been considered by some authors to be generated from the same set of neurones (Jääskeläinen et al., 2004), though subsequent evidence questions these claims (Näätänen, Jacobsen & Winkler, 2005). Later component differences may indicate an impact on later levels of speech processing (e.g. the N400 as a marker of semantic processing and the P600 as a marker for syntactic processing).

Whilst temporal prediction may enhance the perception of acoustic parameters which are relevant for phonological perception, the extent to which it affects memory retrieval is not clear: different levels of speech processing might engage mnemonic components to different degrees. For example, phonological perception requires the retrieval of phonemes from long-term memory, and lexical-semantic access of a word would additionally recruit semantic memory processes. To give another example, syntactic processing not only requires lexical access but also an implicit, domain-specific knowledge about the grammatical rules of language. As proposed by Patel (1998, 2008), it may be the case that these aspects of speech processing are stored in representational networks which are specific to the language domain. However, according to this SSIRH framework, these long-term representational networks also interact with online processing resources - which may be shared between music and speech. In this case, musical rhythm could hypothetically impact on the representational networks via the resource networks of language.

6.3 Clinical Perspectives

To what extent can music-based speech therapies draw upon fundamental research? Whilst the standardisation of speech therapy methods is questionable, and that the tailoring of speech therapy methods to the individual is considered to be one of it's strengths, research can shed light as to the underlying causes of a speech impairment as well as by what mechanisms rhythmic input may influence speech. This in turn may highlight what might be important features of speech therapies.

For instance, the rhythmic aspect of melodic intonation therapy is found to play a more potent role in enhancing speech fluency in non-fluent aphasic patients, compared to the melodic aspect (Stahl et al., 2011, 2013). In this case, not only might speech therapists place a greater focus on the rhythmic aspects of this therapy, but it also supports the idea that rhythmic input can bypass damaged left hemisphere networks which are normally dedicated to speech timing and engage right-hemisphere structures which are also involved in speech production. In speech disfluent populations whose neural correlates is less well-characterised, such as stuttering (Yairi, 2007), it may be more difficult to make conclusions about via what mechanisms musical rhythm might enhance fluency (see Appendix 2).

Studying the perception or production of temporal structures in populations who have a temporal perception or production deficit may also tell us more about perception-action interactions. For instance, the finding that an inability to synchronise with a temporal sequence is coupled with an impaired perception for temporal sequences may demonstrate this link (Phillips-Silver et al., 2011). Indeed, in Appendix 2, we observed a greater priming effect on speech reproduction in PWS who had better rhythm discrimination and rhythm reproduction abilities. This complements findings with unimpaired populations in Experiment 2, whereby subjects (in the group who underwent audiomotor training with the four primes) were better able to reproduce prime 2 and 3 - the two primes for which a rhythmic priming effect on speech perception was significant. In this instance, it may be hypothesised that a greater ability to perceive a rhythm allows it to have a greater priming effect. The link between perception and action of temporal structures has also been noted in clinical populations with movement disorders; Hopyan, Schellenberg & Dennis (2009) found that individuals suffering from spina bifida (a developmental congenital disorder also causing rhythmic movement disorders) also have a deficit in metre perception. Additionally, Conway & Christiansen (2005) found that individuals who have been deprived of auditory input (HI children) do not develop normal rhythmic motor sequencing abilities. These findings could suggest that an impairment of production arises through an impaired perception, and vice versa. However, rhythmic deficits rooted either in action (e.g. Hopyan et al., 2009) or perception (e.g. Conway et al., 2009) may result in a disruption of, or an undeveloped, central timing processes which are important for both perception and production of temporal behaviour. Alternatively, it may be the audio-motor links which are disrupted, which can explain findings that an impaired temporal production (synchronisation) is not necessarily accompanied (i.e caused) by an impaired perception (Sowiński & Dalla Bella, 2013).

Another consideration is that of individual differences, and that not all patients or subjects equate to the 'average' in terms of responsiveness to either a treatment or experimental protocol. This is particularly evident in the experiment conducted with 8 PWS (Appendix 2); in addition, stuttering is known for its variability of symptoms (e.g. some people suffer only from blocks, some from primarily repetitions, some from only facial tics); indeed, it has been proposed that developmental stuttering is closely related, if not a subtype of, Tourettes Syndrome and that there may be many manifestations of stuttering, with perhaps different underlying causes (Yairi, 2007). This can also be observed in differences between neurogenic and developmental stutterers, whereby stuttering caused by brain injury is more characterised by repetitions of entire words and phrases, as opposed to single sounds.

With respect to our study performed with HI children (Experiment 3), there are also individual differences which separate the group. Namely, this difference was between CI users and HA users. We report evidence that children who used CIs (who also suffered from a greater hearing impairment) benefited more from rhythmic priming. Whilst this may be due to the greater room for improvement in the CI group (their baseline articulation was at a lower level to HA users), it may also identify the need for perhaps a more rhythm-based speech therapy, at least compared to children who use HAs. Indeed, aside from the low spectral resolution transmitted by CIs compared to hearing aids and the consequently greater reliance on temporal cues (e.g. Shannon et al., 1995), findings from the literature on temporal orienting can also add to the idea that rhythmic structures may be especially important for CI children. For instance, musical training (in normal hearing people) results in a greater ability to detect speech under conditions of noise (Anderson et al., 2010; Parbery-Clark et al., 2009), which may come about through a greater neuronal entrainment to the temporal aspects of speech (Zion Golumbic et al., 2012). CIs can be especially sensitive to noise, and speech perception must occur in an acoustically-challenging environment in which there is little spectral information. Rhythmic or musical training may enhance the capacity for CI children to extract temporal cues required for speech in much the same way as normal listeners.

6.3.1. Long-term benefits?

The current thesis has investigated the online effects of musical rhythmic priming on speech abilities. However, the long-term benefits have not been addressed - an issue which has priority in a clinical context. Whilst the Experiments conducted aimed to assess the relationships between music and speech temporal processing, findings that HI children can benefit from rhythmic priming might also suggest that a rhythm-based method can enhance the aid speech acquisition in this population (which is obviously considered to be a long-term developmental milestone). Further longitudinal studies are required to assess this claim.

In other cases, most notably in those suffering from speech fluency disorders, the long-term benefits of auditory rhythm are not well-known. It can also be difficult to disentangle a true effect of the speech therapy from spontaneous recovery, e.g. in the case of aphasia due to stroke. Should auditory training have the ability to retrain speech motor programmes (e.g. Thaut, 2005), training with musical rhythm should see a benefit in spontaneous speech outside the context of therapy. Indeed, considering the huge amount of literature presenting evidence for the plastic changes induced by musical training (e.g. those subsiding audio-motor connectivity, Grahn & Rowe, 2009), even after short periods of training (Bangert et al., 2006; Jääskeläinen et al., 2007) alongside evidence that music-induced plasticity can 'transfer' to speech (e.g. of subcortical encoding, Wong et al., 2007), it is an easy claim to make; musical rhythmic training with patient populations might be hypothesised to induce structural changes which are able to benefit speech in the long-term, outside the context of materials used in speech therapy sessions.

In a recent study with left-hemisphere brain damaged non-fluent aphasic patients (Stahl et al., 2013), it was found that rhythm-based treatments (singing/rhythmic speech) were equally similar in enhancing the production of phrases, whose production relies on RH structures. Whilst this both challenges views that singing induces a hemispheric transfer of speech production (left to right hemispheres) and provides evidence that the rhythmic aspect (as opposed to the melodic aspect) of singing is what may drive fluency, this study also compared a music/rhythm-based treatment with standard speech therapy. Contrary to what one might hope, the authors found that whilst music-based therapy more greatly improved performance within the formulaic material used during the sessions (compared to those who underwent standard speech therapy), this did not extend to speech contexts outside that of the therapy sessions, whilst the standard speech therapy did. This suggests that music-based therapies alone may not benefit long-term effects in speech, and thus that processes engaged by music may not apply to processes required by speech. Similarly, whilst PWS are more able to produce fluent speech whilst speaking alongside a metronome/other speakers compared to producing solo speech (e.g. Toyomura et al., 2011), the ability for this to apply to real-life speaking situations - in which there is no external cue - has not yet been supported. In one case with (Appendix 2), it was observed that whilst one stuttering subject (a musician) claimed he was not able to provide himself with a tapping cue (which, when provided by someone else, greatly enhanced speech fluency), he was however able to improve his speech fluency in non-rhythmic trials by "imagining the rhythm". The role of musical rhythm imagery on speech fluency (as a substitute for an external cue) has yet to be investigated further; though it should also be noted that imagery would also be considered to provide an online 'cue' and would not necessarily address any questions regarding the long-term effects on speech motor networks, but rather the long-term effects on the ability to imagine an external cue.

To conclude, whilst we have measured the effect of engaging online temporal processes on speech, evidence that auditory rhythmic training may benefit speech production in the long-term is lacking, variable or anecdotal. The potential for long-term benefits are also likely to be highly dependent on the speech impaired population in question. In the case of HI children in which hearing loss can cause an articulatory disorder, rhythmic primes or training may aid speech acquisition (which is certainly a long-term effect) whilst in those suffering from fluency disorders, further research is required to assess the long-term benefits of music-based speech therapies, or whether music-induced plasticity after a period of training could impact on speech fluency.

6.3.2. Auditory vs. multimodal rhythmic engagement

For final consideration is that of multi-modal rhythmic engagement. Multimodal training is thought by many to be more efficient form of therapy, not only in speech (e.g. Guberina, 1963), but also in motor rehabilitation (e.g. Maegele et al., 2005). The Verbotonal Method (Guberina, 1963, 1981; Asp, 1985) is a speech therapy method proposed for HI children. This method combines rhythmic movement, sound and tactile stimulation in order to improve spoken language. At least in terms of combined audio-motor training, there is evidence to suggest this can influence perception of temporal structures. This improved perception may also impact on production abiities.

The effect of combined audiomotor training with musical rhythm has been found in several studies to influence the perception of temporal structures; movement to ambiguous metrical sequences influences how it is subjectively grouped in both infants (Phillips-Silver & Trainor, 2005) and adults (Phillips-Silver & Trainor, 2007), motor engagement with rhythm enhances beat extraction (Su & Pöppel, 2012) and musical training enhances sensitivity to metrical hierarchies (Geiser, Sandmann, Jäncke & Meyer, 2010). A recent study by Manning & Schultz (2013) also shows that motor engagement (tapping) with an auditory beat improves the ability to detect whether a final, target tone occurs on-beat or later than expected. In other words, motor engagement with the auditory rhythm is thought to enhance the potential for perceptual rhythmic priming. Together, these results demonstrate the role of audiomotor engagement in timekeeping, beat perception, and metrical perception/sensitivity. It may be that this is because temporal processes are subserved by a central timekeeper (i.e. domain- and modality-general temporal processes), in which case multimodal engagement during speech therapy should be most beneficial as these processes are more greatly engaged, and sensorimotor connectivity enhanced.

In Experiment 2, we found that motor engagement with a rhythm (rather than passively listening to it) enhances it's propensity to impact upon speech perception, even after only a short period of familiarisation (in this case, 20 repetitions for each rhythmic prime); even short periods of training can induce plasticity (e.g. Jääskeläinen et al., 2007, review Herholz & Zatorre, 2012). However, it should be noted that, rather than rhythmic training having a facilitatory effect on speech processing, we instead found that a mismatching prime-sentence structure resulted in a delayed processing compared to matching conditions. Whilst this may suggest that audiomotor training may not be beneficial in rhythmic priming, these results nonetheless reveal an impact on rhythmic training (short-term) on the processing of linguistic metre in relation to musical metre. A longer period of training may be expected to augment this effect and perhaps to reveal an advantage of 'matching' conditions in phonological processing. Also, these subjects were speech unimpaired and it may be the case that audiomotor training may be beneficial in speech pathologies; indeed, we adopted this in Experiment 3.

6.4. Future perspectives

The research conducted during this thesis has measured the impact of rhythmic priming on phonological processing. Behavioural (reaction time and articulation performance) and EEG evidence has been presented. From a theoretical perspective, this suggests that music and speech may share temporal processes; the fact that musical temporal structures has the capacity to impact on speech processes suggests that these processes are domain-general, and evidence exists to suggest that this is perhaps also modality-general. Whilst we have investigated the impact on phonological processing, rhythmic priming may also enhance other linguistic functions, such as syntax (Przybylski et al., 2013).The relative contributions of different temporal structures (i.e. beat/metre/durational variability) as well as the potential temporal processes engaged (e.g. dynamic attending/abstract representations) to different levels of speech processing has yet to be understood.

From a more clinical perspective, the findings presented highlight the reliance of temporal cues in CI users and suggests that rhythm could be beneficial for the development of speech articulation abilities. Whilst we hypothesised that the enhanced articulatory performance found for metrically predictable contexts was a result of an enhanced perception of the target sentence, this also requires further research. Electrophysiology might be used to measure the processing of target sentences in predictable/unpredictable contexts. As to two possibilities, cABRs are commonly used to measure the efficiency with which CI users are able to encode sound and cortical AEPs are also used to measure the efficiency of synaptic transmission of auditory information to cortical areas (e.g. Sharma & Dorman, 2006).

The use of musical rhythm as a prime, as opposed to a speech-based prime might also be further investigated. Here, a musical rhythmic prime (which presented listeners with the rhythmic context) was chosen due to music's strong ability to induce perceptual and motor synchronisation. However, it may also be the case that a spoken sentence (perhaps without segmental information and displaying the same temporal regularity as musical temporal structures) would just as well prime for the temporal structure: despite the fact speech does not display the same degree of temporal regularity as music under normal speaking conditions, one view is that temporally regular speech has just as much potential to induce synchronisation as temporally regular musical rhythm (Dalla Bella et al., 2013; Lidji et al., 2011a), and may engage the same attentional mechanisms as music (Arantes & Barbosa, 2010). In this case, it may be interesting to compare the priming ability of speech and musical rhythm to ascertain exactly how much more of an advantage musical rhythm might actually afford. In this case, it can be predicted that a speech prime which is temporally equal to a musical prime would similarly enhance the processing of a speech target. There may also be additive effects from the melodic aspects of music, which have not be considered here.

Whilst rhythm may have an enhancing effect on language abilities in speech-impaired populations, the mechanisms for this may differ depending on the speech disorder. For instance, we suppose that articulation improvement occurs via an enhanced perception in HI children (i.e. children with both speech perception and production impairment). Those suffering from more motor-based speech impairments, such as stuttering, might also benefit from rhythmic cues, but rather through the motor influence of rhythm on production (i.e. speech production is impaired).

The three experiments of this thesis were conducted in French. Under the supposition that speech engages similar temporal processes regardless of the language, rhythmic priming would be hypothesised to be effective across languages, and via similar mechanisms: whilst the temporal structure representations (long-term) would be expected to differ depending on native language (since each language has its characteristic rhythms which are implicitly learned), the online processes involved in encoding temporal structures (dynamic allocation of attention/abstract representation) would not be expected to differ. However, it is possible that a more metrically regular (e.g. 'syllable-timed') language might engage attentional processes to a greater degree, in the same way that artificial speech metres which cannot be strictly considered as 'natural' might (for example, those constructed for experimental purposes, or those present in poetry).

Overall, future research might consider how other levels of speech processing might benefit from rhythmic priming, the relative contributions of beat and metre, or of dynamic attending and abstract representations, the mechanisms via which speech articulation might be enhanced (i.e. via an enhanced perception, or not), the advantages of using multimodal vs. auditory-only rhythmic input, whether musical rhythm has a greater priming effect than speech rhythm, whether there are long-term benefits of rhythmic training on speech abilities, and the effect of rhythm in different speech-impaired populations, or in different language contexts.

APPENDIX 1

One theme of this thesis was to investigate to what extent music and speech may similarly draw upon temporal processes. In the Experiments reported, we have investigated this by measuring the online impact of musical rhythm on speech perception and production. Should there indeed be shared temporal resources which are more representational, we should perhaps be able to observe an effect of linguistic rhythm on music.

There is evidence that experience within a native language can impact on musical rhythm perception: the way we segment speech strings, according to rules of our native language, also seem to govern our segmentation of rhythmical patterns into rhythmic groups (rhythmic grouping). In order to study the effect of linguistic rhythmic representations on musical rhythm production, we conducted a behavioural study with French and English native speakers.

Report: The effect of native language on musical temporal structures

Introduction

Leaving aside the categorisation of languages into rhythmic classes, languages do display their own typical rhythmic characteristics. For instance, whilst English words present both iambic and trochaic stress patterns and have variable inter-vocalic intervals, French words consistently present iambic patterns and is less variable inter-vocalic intervals. The fact that speech comprehension is dependent on these implicit ‘rules’ - the way in which stressed and unstressed syllables are organised in time - is the aspect in which speech can be considered to display temporal predictability. Importantly, this is dependent on experience within a particular language environment.

There are several lines of evidence which indicate an effect of native language on perception and production of musical rhythm. It has been found that the durational variability in French and English music reflects that of their speech (Huron & Ollen, 2003; Patel & Daniele, 2003) (Section 2.3). Additionally, listeners are able to classify French and English instrumental songs according to the language of origin, presumably through the rhythmic characteristics (Hannon, 2009). Native language has also been found to effect rhythm perception. Despite previous work, suggesting that universal principles govern rhythmic grouping into 'loud-soft' and 'short-long' rhythmic groups (Bolton, 1894; Woodrow, 1909; Hay & Diehl, 2007), research has found that rhythmic characteristics of one’s native language influences the way in which musical rhythm is grouped, supposedly according to the implicit rules of speech segmentation (Kusumoto & Moreton, 1997; Iversen et al., 2005). These results indicate that implicit knowledge (long-term representations) about our native language rhythm influences our perception and production of rhythmic structures in music. We hypothesised that 1) musical rhythms would be more easily reproduced when they represented rhythms of a participants native language, and 2) these rhythms would also be better implicitly learned than rhythms derived from non-native language.

In order to test this hypothesis, we designed a behavioural experiment carried out by a native French-speaking group (F Group) and native English-speaking group (E Group). Musical rhythmic sequences (derived from French and English sentences) were heard and reproduced by both groups, with the prediction that participants would better reproduce musical rhythms of their own native language, due to the fact they were more familiar with their perceptual and motor associations. Participants then performed an implicit learning task in which they had to decide whether rhythms were ‘familiar’ or ‘unfamiliar’, with the prediction that listeners would better retain musical rhythms of their native language.

Methods

Stimuli

20 French and 20 English sentences were taken from the experiment by Nazzi et al. (1998). These sentences have been previously used in other studies investigating musical and speech rhythm (Patel & Daniele, 2003, 2006). Since the intention was to convert these sentences to rhythmic stimuli, sentences were reduced to a more feasible length of 8 - 12 syllables. The resulting French sentences had a mean syllable length of 10.1 (s.d. 1.2) and English sentences had a mean syllable length of 10.3 (s.d. 1.3). The resulting sentences were then read aloud by a bilingual French/English speaker, who, instead of reading the full sentences, used only the sound ‘ta’, whilst retaining the prosody of the sentence. For instance, the word “difficult” would have been spoken as “ta ta ta”, with accentuation placed on the first and last ‘ta’’s. ‘Ta’ sequences were transcribed into musical rhythms by two musicians (one English, one French). It might be expected that the French and English transcibers would interpret the ‘ta’ sequences differently. However, both were in agreement - possibly since they were to a larger range of rhythmic variability in music than might a non-musician (both were experienced jazz musicians).

Resulting English rhythms were significantly more variable than french rhythms, as measured by the nPVI (independent samples t-test, p= 0.0014). This concurs with the literature (Patel & Daniele, 2003). French-derived rhythms had a mean nPVI of 29.99, s.d. 7.29, and English-derived rhythms had a mean nPVI of 37.19, s.d. 15.8. Transcriptions were then converted into audio files. The base duration (corresponding to a semi-quaver) was 200 msecs, around the same duration of a spoken syllable.

Participants

15 native french speakers and 15 native English speakers took part in the study. The F-Group was aged 18 - 25 years (mean = 21.07, s.d. = 1.68), The E-Group was aged 19 - 25 (mean = 22.13, s.d. = 1.5). 10 subjects in the E-Group and 7 subjects in the F-group had had musical training but none considered themselves to be professional musicians.

Protocol

In the rhythm reproduction task, participants were presented with 15 French-derived and 15 English-derived rhythms, which were randomised and counterbalanced for order. Participants were asked to reproduce the rhythms using the sounds "ti" and "ta". Recordings were made of their performance.

In the IL task, participants heard 20 rhythms (10 English-derived, 10 French derived; 10 new rhythms, 10 old rhythms) and were asked to decide whether the rhythm they heard was 'familiar' or 'unfamiliar' (i.e. whether it had been present in the rhythm reproduction task).

Scoring and Analyses

Rhythm reproduction was scored out of 9 by two people.

Implicit learning was scored as a d-prime score, which measures the ability to correctly identify 'familiar' items.

For rhythm reproduction, we ran a 2 x 2 ANOVA with the factors 'Group' (F/E) and 'Rhythm type' (French-derived rhythm / English-derived rhythm). For IL, we similarly ran a 2 x 2 ANOVA with the factors 'Group' (F/E) and 'Rhythm type' (French-derived rhythm / English-derived rhythm).

Finally, we also ran correlation with the factors 'Reproduction of French rhythms', 'Reproduction of English rhythms', 'IL d-prime score for French sequences', 'IL d-prime score for English sequences', 'Group' (English/French), 'IL% correct overall' and 'Rhythm discrimination' as measured by the Musical Ear Test (MET) (Wallentin et al., 2010).

Results

1. Rhythm reproduction

There was a main effect of rhythm: French rhythms were more easily reproduced than English rhythms (F(1,27)=21.858, p=0.00007). There was no Group x Rhythm Type interaction; both English and french participants performed similarly (F(1,27)=0.519, p=0.477).

Rhythm reproduction scores were positively correlated the MET score (p<0.05, values 0.58 & 0.61).

2. Implicit learning

Both groups retained similarly French and English rhythms; no effects were observed for the IL task (F(1, 27)=0.665, p=0.42), and there was no interaction of Rhythm Type with Group (F(1, 27)=0.0142, p=0.905).

Discussion

We hypothesised that musical rhythms derived from one's native language would be both better reproduced and implicitly recognised. This is based on research showing that French and English music well represents French and English speech, in terms of durational variability (Patel & Daniele, 2003; Huron & Ollen, 2003).

We hypothesised that, due to the motor familiarity with, as well as the perceptual familiarity with the durational characteristics of ones native language, participants would show a preference for musical rhythms derived from their native language, in terms of both reproduction and implicit learning. Alternatively, we hypothesised that English speakers, with more experience within a language displaying higher durational variability, would outperform French speakers in English rhythms, and perform equally well in (less variable) French rhythms.

There are several reasons as to why we did not find this to be the case. Firstly, many of the subjects in the current study had had musical training. Secondly, the tasks were difficult. Thirdly, the rhythms used in the experiment relied upon a speaker-based knowledge which was not explicitly (or implicitly) related to speech. Fourth, not all participants were entirely naive to the respective foreign language (English/French). Fifth are considerations as to the scoring methods we used. Finally, it remains possible that native language does not impact on musical rhythm production. Whilst participants may have perceived or grouped the rhythmical sequences differently, this may not be reflected in production, for instance.

The overlap which is thought to exist between speech and music temporal processing theoretically allows for a bidirectional influence: not only can music impact on speech perception (as described in the introduction), but speech can also impact upon music perception. This impact of native language would also be expected to extend to rhythm production. However, the studies we have conducted which measure this use a method of rhythmic priming, and have also not investigated the role of durational patterning (as has this study). The rhythms used in the current study probably draws upon different resources (e.g. greater referent memory processes). It may be the case, therefore that priming with speech (to make an explicit connection between the speech and the music) may impact on rhythm reproduction.

APPENDIX 2

Report: Rhythm and speech fluency in People who stutter (PWS)

Introduction

Chronic developmental stuttering is a speech-specific motor disorder thought to be caused by a disruption of the timing of speech processes. Stuttered speech is characterised by a disrupted speech prosody. This can include involuntary sound repetitions, prolongation of sounds, a difficulty to initiate speech ('blocking'), and may also be accompanied by facial tics (Ludlow & Loucks, 2004; Yairi & Ambrose, 1999). A stutter can present itself differently in a given individual, however, and its development is also variable (Bloodstein, 1960; Jiang et al., 2012). Whilst the onset of stuttering tends to be at around 3-6 years of age, spontaneous recovery occurs in some 75 - 80% at adolesence (Bloodstein, 1993; Mansson, 2000; Yairi & Ambrose, 1999, 2005); it is in those individuals who continue to experience stuttering into adulthood (about 1% of the population) that the term 'chronic developmental stuttering' applies. As of yet there is no neurological explanation as to why some children and not others experience spontaneous recovery (Yairi et al., 1996), though girls are more likely to recover than boys (Yairi & Ambrose, 1999); males are more likely to experience stuttering in adulthood with the male-to-female ratio at 4:1 or 5:1 (Andrew & Harris, 1964; Bloodstein, 1995; Yairi and Ambrose, 2005). The view that stuttering is largely hereditory (Howie, 1981; Kidd, 1984; Yairi et al., 1996) may also explain the higher incidence of stuttering in males (Suresh et al., 2006).

The neural mechanisms underlying developmental stuttering are still under debate, though in PWS it is broadly agreed that it results from a disruption in speech motor control. Specifically, it has been suggested that developmental stuttering is a biproduct of abnormal sensorimotor feedback and audio-motor integration (Caruso, 1991; Max et al., 2004). Max et al., (2004) propose that there is a misalignment between incoming auditory information (self-made speech) and the motor act of speaking. Supposedly, this audio-motor misalignment results in a reissuing of the motor command, and this results in a stutter. The BG, structures known to be important for speech motor automacity (Booth et al., 2007; Wildgruber, Ackermann & Grodd, 2001) are also implicated as structures underlying this speech disorder (Alms, 2004; Giraud et al., 2008; Lu et al., 2010).

There are several linguistic aspects which impact on stuttering frequency (review: Karniol 1993/5). For instance, increased cognitive demands increase disfluencies: in sentences rather than words, in spontaneous, unplanned (rather than read) speech (Wingate, 1988), at the onset of / during syntactically more complex sentences (Wall et al., 1981), at sentence-initial positions (Barr, 2001, review), at sentence-initial positions when the sentence which follows is longer (Karniol, 1995; Oviatt 1995), when the topic is unfamiliar (Merlo & mansur, 2004), at less frequent words (Hubbard 1998), and at stressed syllables (Wingate, 1976, 1988; Prins, Hubbard and Krause, 1991). PWS also exhibit semantic processing deficits (Bosshardt, 2006). To summarise, stuttering is a speech fluency disorder which appears to be sensitive to linguistic complexity and which manifests itself as a speech motor timing disorder, as reflected by abnormal brain activations in areas involved in speech planning, speech production and audiomotor integration. It has yet to be seen in what way linguistic and motor aspects of the disorder interact, or whether these are two symptoms of an underlying timing deficit.

Rhythm as a potential tool in speech therapy for PWS

The use of rhythm for enhancing speech fluency in PWS has been recorded from the 1960's. Early work found that when PWS spoke alongside a metronome beat, they produced less stutters when the beat was rhythmic (Fransella and Beech, 1965), a finding which has been reproduced (Brayton & Contour, 1978; Brady, 1971). Regularising prosodic patterns in speech can also enhance speech fluency in PWS (Andrews et al, 1980, 1983; Karniol, 1995; Wingate, 1988; Packman, Onslow & Menzies, 2000). For instance, 'Syllable-timed' speech training, a common behavioural treatment for stuttering whereby each syllable is produced in time to a beat and with equal prosodic stress, seems to be an effective long-term treatment for children who stutter (Coppola and Yairi, 1982; Trajkovski et al, 2009). There are other ways to produce rhythmic speech aside from 'uniformity' (1 syllable per beat), such as speaking along in time to a beat (Ingham 1984, book). For instance, stressed syllables may align with strong beats as in 'patterned rhythmic speech cueing', described by Thaut (2005). Praying, speaking in unison and speaking in a foreign accent can also enhance fluency, and finally, singing, inherently rhythmical, results in fluent speech in PWS (Giraud et al, 2008; Kell et al., 2009; Webster, 1980; Healey et al, 1976; Andrews et al, 1982; Colcord and Adams, 1979; Davidow et al., 2009). The rhythmic aspect of singing has been found to be key in enhancing speech fluency in other populations with basal ganglia abnormalities (Stahl et al., 2011, 2013).

Auditory rhythm has the ability to retrain motor programmes (Thaut, 2005) and sensorimotor synchronisation is able, to some extent, to induce long-term plastic changes in the brain (see Jeffries et al., 2003; Preibisch et al., 2003; De Nil et al., 2001; 2003; Neumann et al., 2003; Preibisch, et al., 2003). Brain activity of PWS is said to be normalised through ‘fluency shaping’ therapies, a significant component of which is rhythmicity of speech (Giraud et al., 2008; Kell et al., 2009; Webster et al., 1980; Toyomura, Fujii and Kuriki, 2011). In addition, Max & Yudman (2003) found that speech and non-speech mouth movements were also unimpaired when they were produced isochronously.

Rhythmic production is unimpaired in PWS (both continuation and synchronisation tapping) (Max & Yudman, 2003). One view is that the motor timing deficit is therefore speech-specific (Zelaznik et al., 2004). Since one hypothesis for the underlying mechanisms of stuttering may be audio-motor asynchrony (Max et al., 2005), it may be that auditory rhythm, which is intimitely tied to motor synchronisation, can ameliorate speech timing via a similar mechanism observed in patients with other motor disorders (review Thaut & Abiru, 2010; Thaut et al., 1999). Rhythm-based therapies may capitalise on strong audiomotor connectivity, which is reflected by spontaneous synchronisation with auditory rhythm.

We investigated the effect of auditory metrical priming on sentence production. The task was to reproduce heard sentences. A matching or mismatching rhythmical prime or a non-rhythmical control prime followed the heard sentence. There were two types of sentence rhythm: regular’ (trochee trochee trochee trochee stress pattern) or ‘irregular’ (trochee iamb iamb trochee). We took measures of syllable, consonant and word accuracy (%), speech onset time, and the duration of blocking. We hypothesised that speech would be more fluent when sentence reproduction was preceded by a matching prime.

Methods

Participants

9 PWS, all males, took part in the behavioural experiment. One participant was excluded from further analyses as he appeared to suffer from neurogenic stuttering rather than developmental stuttering (stuttering manifested itself as whole word/sentence repetition, he had suffered from a serious head injury and the onset of his stutter was shortly after this). This left 8 PWS. Participants had an average age of ***, * were not currently in therapy and the number of years spent in speech therapy ranged from ** to ** years. Musical expertise was variable. Participants were recruited through a poster advertisement in collaboration with Speech and Brain Lab, Oxford.

Design & Materials

Asesesement of Stuttering: Stuttering Severity Instrument 3 (SSI-3): two speech samples were read aloud. Inclusion criteria was 5% or more fluency errors in the reading of this passage which classified participants as PWS. (Riley, 1994)

Experimental Task:

Participants heard a target sentence followed by a prime which either matched or mismatched the metrical structure of the sentence. Their task was to reproduce the sentence once the prime had ended. For this, there were two blocks: 1 'simple' rhythm block and one 'complex' rhythm block. Simple rhythms had a X x X x . . X x X x . . structure (matched by sentences with a trochee, trochee trochee trochee structure), and complex rhythms had a X x x X . x X . . X x structure (matched by sentence with a trochee iamb iamb trochee structure). A control block consisting of a non-rhythmical prime (variable ISI durations and mismatching stress patterns, but which still contained four strong and four weak prime beats). Percussion sounds were used to create these 3 primes.

Sentences were all 4 words long and contained 8 syllables, and were created by Hubbard (1998) in a study which investigated how stress patterns may impact on stuttering frequency. The corpus contains 20 sentences with the same rhythm as the 'simple' prime, and 20 sentences with the same rhythm as the 'complex' rhythmic prime. Sentences were matched for word frequency and had low 'meaning' (syntactically correct but semantically unlikely). There were 13 trials in each block. Block order ('simple' prime block, 'complex' prime block, control block) and sentence condition was randomised across participants. Sentences were recorded by a professional voice actor for the purposes of this study.

Participants were instructed to listen to the sentence, listen to the prime and then reproduce the sentence. No explicit information regarding the relationship of the sentence to the prime was provided.

Rhythm Production:

After experimental trials, participants were asked to reproduce the two rhythm heard in the experiment using the mouth (sounds 'ta' and 'ti' given as an example but participants chose their preferred sounds).

Rhythm perception:

Participants performed the rhythmic perception section of the Musical Ear Test (Wallentin et al., 2010). Participants heard two consecutive rhythms and had to decide whether they were the same or different. There are 52 trials in this test, scores were converted into percentages.

Questionnaire on therapy and musical background:

Participants provided information about their musical background and musical listening experience as well as their therapy history.

Analyses / scoring

1. Experimental Task:

Speech of PWS was scored for accuracy at the level of the syllable (fluency out of 8), consonant (out of the no. of voiced consonants present in the sentence) and word (out of 4). The absence of stuttering-specific disfluencies (blocking, repetitions, unusually accented sounds, lengthened sounds) would result in an accuracy score of 100%. Speech onset time, position of disfluency within the sentence and blocking/hesitation duration were also measured.

2. Rhythm Production

Reproduction of rhythmic sequences was scored out of 9.

Results / Observations

Experimental Task:

% Fluency: There was no effect of prime block (‘simple’/‘complex) on % fluency at the syllable, consonant or word level (RM ANOVA, block (2) x measure (3)). There was a significant difference between matching and mismatching conditions (mismatching higher fluency % than matching conditions (t-test, p=0.048), though a Wilcoxon test is more suitable for this data, and this resulted in a no significant differences between the two conditions. When including the control block in the analyses, there was a trend for control and match conditions to be at a lower fluency than mismatching conditions at the word level only (p=0.08, RM ANOVA, condition (3) x measure (3)).

Blocking duration: Though not significant, it seemed that blocking duration and variability also followed this pattern.

Speech onset: Mean speech onset, no significant differences for either block (‘simple’ prime 1, ‘complex’ prime 2, control) or condition (match/mismatch/control). Distribution of speech onset (observations: not enough data for statistical analyses). The prime seemed to impact on speech onset time and occurred ‘on-beat’. For ‘simple’ prime 1, the beat level is 400 msec and most onset times seemed to occur at 1.2 secs (3 beats) after the final beat of the prime end as well as smaller peaks at 1.6 secs (4 beats after) and 2 secs (5 beats after), and seemed also earlier to the control. For ‘complex’ prime 2 which had a beat level of 600 msecs, speech onset was more common at 600 msecs (1 beat after) and 1.2 secs (2 beats after). There seems to be no differences between match and mismatch conditions for this distribution. There are not enough observations for statistical analyses.

Position of disfluency: Disfluencies most likely to occur on word onsets (no statistics).

Considerations:

Instructions

Participants were not explicitly instructed to use the rhythmic prime as a template for their speech production.

Analyses

Speech in fluent speakers in never 100% fluent (for example contains hesitations and incorrect phonological production). Isolating stuttering-specific disfluencies (blocking, repetitions, unusually accented sounds, lengthened sounds) from normal disfluencies has its drawbacks since there is no telling to what extent stuttering itself is responsible for these ‘normal’ disfluencies. For instance, PWS commonly replace words in order to avoid stuttering which would result in a phonological production error. However, the analyses was also run on ‘total’ disfluencies (incorrect phonological production, self-corrections, and presence of ‘um’/‘ah’s) and similar results were obtained. Speech fluency was measured subjectively. Rhythm reproduction was also assessed subjectively. However, there has been shown to be coherence between subjective and objective measures in assessing vocal (melodic) accuracy (Larrouy-Maestri et al., 2012). Also, there were very few trials (13) per block. These small numbers were not compensated for by a large sample size making statistical analyses difficult.

Rhythmic abilities

PWS do not have timing problems in producing isochronous movements in either speech or non-speech (Max & Yudman, 2003), but the processing of meter is distinct to that of beat. It may be the case that entrainment to a regular beat is able to enhance speech fluency, but that additional metrical information is not useful. Also, rhythm production was measured after the experiment. In participants in which rhythm production was poor, it seems unlikely that they would have benefited from matching conditions in the experimental trials. For this reason, providing rhythmic training prior to the experiment may have made participants more aware of the rhythms which they heard during the experiment, and thus may have enhanced any experimental effect. Finally, participants with poor rhythm production and perception abilities were not discarded from any analyses. Perhaps an inclusion criteria of ‘normal’ rhythm perception/production abilities should have been implemented.

Individual Differences

Participants varied greatly in their severity of stuttering (measured by SSI-3), the presentation of their stutter (e.g. some participants did not block but produced many repetitions, and vice versa), their responsiveness to rhythmic priming and their rhythm perception and production abilities. For instance, the most severe stutter (of MS08) was reduced during rhythmic conditions vs. arrhythmic control condition (below, y axis % accuracy), and when external tapping (beat) was provided (during interviews/speaking outside the experiment) this participant was able to produce fluent speech. He suffered from severe blocks but rarely presented repetitions. All other participants suffered from mild - average stuttering disfluencies.

For now developmental stuttering is regarded as a single thing. With more participants, perhaps subgroups of stuttering may have been possible to identify which would predict their response to rhythmic priming (Yairi, 2007). It is important to bear in mind that speech therapy is tailored to the client since individual differences are so great. There is probably no advantage to the standardisation of speech therapy methods. Whilst some general conclusions might be drawn from this study (e.g. observation of speech onset distribution), the main acknowledgement here should be the huge variability of PWS and that the flexibility of speech therapy itself is probably what allows it to be successful, and also that 8 subjects is not enough for any conclusions to be drawn.

REFERENCES

Abercrombie, D. 1967: Elements of general phonetics. Edinburgh: Edinburgh University Press.

Abraham, G. (1974). The tradition of Western music (pp. 62–83). Berkeley, CA: University of California Press.

Ackermann, H., Gräber, S., Hertrich, I., & Daum, I. (1997). Categorical speech perception in cerebellar disorders. Brain and language, 60(2), 323-331.

Aksentijevi, A., Elliott, M. A., & Barber, P. J. (2001). Dynamics of perceptual grouping: Similarities in the organization of visual and auditory groups. Visual Cognition, 8(3-5), 349-358.

Albert, M. L., & Bear, D. (1974). Time to Understand A Case Study of Word Deafness with Reference to The Role of Time in Auditory Comprehension. Brain, 97(2), 373-384.

Albert, M. L., Sparks, R. W., & Helm, N. A. (1973). Melodic intonation therapy for aphasia. Archives of Neurology, 29(2), 130.

Allen, G. D. (1972). The location of rhythmic stress beats in English: An experimental study I. Language and Speech, 15(1), 72-100.

Alm, P. A. (2004). Stuttering and the basal ganglia circuits: a critical review of possible relations. Journal of communication disorders, 37(4), 325-369.

Anderson, S., Skoe, E., Chandrasekaran, B., & Kraus, N. (2010). Neural timing is linked to speech perception in noise. The Journal of Neuroscience, 30(14), 4922-4926.

Andrews, G., Craig, A., Feyer, A.-M., Hoddinott, S., Howie, P., &Neil- son, M. (1983). Stuttering: A review of research findings and theories circa 1982. Journal of Speech and Hearing Disorders, 48, 226-246.

Andrews, G., Guitar, B., & Howie, P. (1980). Meta-analysis of the effects of stuttering treatment. Journal of Speech and Hearing Research, 45, 287-307.

Andrews, G., & Harris, M. (1964). The Syndrome of Stuttering. London, UK: William Heinemann Medical Books.

Andrews, G., Howie, P. M., Dozsa, M., & Guitar, B. E. (1982). Stuttering: Speech pattern characteristics under fluency-inducing conditions. Journal of Speech, Language and Hearing Research, 25(2), 208.

Arantes, P., & Barbosa, P. A. (2010, May). Production–perception entrainment in speech rhythm. In Proceedings of the speech prosody (pp. 1-4).

Arnfield, S., 1994. Prosody and syntax in corpus based analysis of spoken English. Ph.D. dissertation, University of Leeds.

Aschersleben, G. (2002). Temporal control of movements in sensorimotor synchronization. Brain and cognition, 48(1), 66-79.

Asp, C. W. (1985). The verbotonal method for management of young, hearing-impaired children. Ear and hearing, 6(1), 39-42.

Astheimer, L. B., & Sanders, L. D. (2011). Predictability affects early perceptual processing of word onsets in continuous speech. Neuropsychologia, 49(12), 3512-3516.

Astésano, C., Besson, M., & Alter, K. (2004). Brain potentials during semantic and prosodic processing in French. Cognitive Brain Research, 18(2), 172-184.

Ayotte, J., Peretz, I., Rousseau, I., Bard, C., & Bojanowski, M. (2000). Patterns of music agnosia associated with middle cerebral artery infarcts. Brain, 123(9), 1926-1938.

Bangert, M., Peschel, T., Schlaug, G., Rotte, M., Drescher, D., Hinrichs, H., ... & Altenmüller, E. (2006). Shared networks for auditory and motor processing in professional pianists: evidence from fMRI conjunction. Neuroimage, 30(3), 917-926.

Barbosa, P. (2007). From syntax to acoustic duration: A dynamical model of speech rhythm production. Speech Communication, 49(9), 725-742.

Barbosa, P. A., "Explaining cross-linguistic rhythmic variability via a coupled-oscillator model of rhythm production", Speech Prosody 2002 Conference, 163-166, 2002.

Barbosa, P. A., Arantes, P., Meireles, A. R., & Vieira, J. M. (2005, September). Abstractness in speech-metronome synchronisation: P-centres as cyclic attractors. In INTERSPEECH (pp. 1441-1444).

Baroni, M., Maguire, S., and Drabkin, W. (1983). The Concept of Musical Grammar. Music Analysis, 2:175–208.

Barr, D. J. (2001). Trouble in mind: Paralinguistic indices of effort and uncertainty in communication. In C. Cavé, I. Guaïtella, & S. Santi (Eds.), Oralité et gestualité: Interactions et comportements multimodaux dans la communication (pp. 597–600). Paris: L'Harmattan.

Baumann, S., Koeneke, S., Meyer, M., Lutz, K., & Jäncke, L. (2005). A Network for Sensory- Motor Integration. Annals of the New York Academy of Sciences, 1060(1), 186-188.

Baumann, S., Koeneke, S., Schmidt, C. F., Meyer, M., Lutz, K., & Jancke, L. (2007). A network for audio–motor coordination in skilled pianists and non-musicians. Brain research, 1161, 65-78.

Bausenhart, K. M., Rolke, B., & Ulrich, R. (2008). Temporal preparation improves temporal resolution: Evidence from constant foreperiods. Perception & Psychophysics, 70(8), 1504-1514.

Beach, C. M. (1991). The interpretation of prosodic patterns at points of syntactic structure ambiguity: Evidence for cue trading relations. Journal of memory and language, 30(6), 644-663.

Becker, J. B. (Ed.). (2002). Behavioral endocrinology. The MIT Press.

Belin, P., Zilbovicius, M., Crozier, S., Thivard, L., Fontaine, A., Masure, M. C., & Samson, Y. (1998). Lateralization of speech and auditory temporal processing. Journal of cognitive neuroscience, 10(4), 536-540.

Bengtsson, S. L., Ullen, F., Henrik Ehrsson, H., Hashimoto, T., Kito, T., Naito, E., ... & Sadato, N. (2009). Listening to rhythms activates motor and premotor cortices. cortex, 45(1), 62-71.

Besson, M., & Schön, D. (2001). Comparison between language and music. Annals of the New York Academy of Sciences, 930(1), 232-258.

Besson, M., Chobert, J., & Marie, C. (2011). Transfer of training between music and speech: common processing, attention, and memory. Frontiers in psychology, 2.

Bharucha, J. J., & Stoeckig, K. (1986). Reaction time and musical expectancy: priming of chords. Journal of Experimental Psychology: Human Perception and Performance, 12(4), 403.

Bigand, E., & Poulin-Charronnat, B. (2006). Are we "experienced listeners"? A review of the musical capacities that do not depend on formal musical training. Cognition, 100(1), 100-130.

Bishop, D. V., North, T., & Donlan, C. (1995). Genetic basis of specific language impairment: Evidence from a twin study. Developmental Medicine & Child Neurology, 37(1), 56-71.

Bloodstein, O. (1960). The development of stuttering: I. Changes in nine basic features. Journal of Speech and Hearing Disorders, 25(3), 219.

Bloodstein, O., & Bloodstein, O. (1993). Stuttering: The search for a cause and cure. Allyn and Bacon.

Bloodstein, O., & Ratner, N. (1995). A handbook on stuttering Singular. San Diego, CA.

Bohn, K., Knaus, J., Wiese, R., & Domahs, U. (2013). The influence of rhythmic (ir) regularities on speech processing: Evidence from an ERP study on German phrases. Neuropsychologia.

Bolger, D., Trost, W., & Schön, D. (2013). Rhythm implicitly affects temporal orienting of attention across modalities. Acta psychologica, 142(2), 238-244.

Bolger, D., Coull, J., & Schön, D. (in press). Metrical rhythm implicitly orients attention in time, as indexed by improved target detection and left inferior parietal activation.

Bolinger, D. (1978). Intonation across languages. Universals of human language, 2, 471-524.

Bolton, T. (1894). Rhythm. American Journal of Psychology, 6, 145-238.

Boltz, M., & Jones, M. R. (1986). Does rule recursion make melodies easier to reproduce? If not, what does?. Cognitive Psychology, 18(4), 389-431.

Bookheimer, S. (2002). Functional MRI of language: new approaches to understanding the cortical organization of semantic processing. Annual review of neuroscience, 25(1), 151-188.

Booth, J. R., Wood, L., Lu, D., Houk, J. C., & Bitan, T. (2007). The role of the basal ganglia and cerebellum in language processing. Brain research, 1133, 136-144.

Bosnyak, D. J., Eaton, R. A., & Roberts, L. E. (2004). Distributed auditory cortical representations are modified when non-musicians are trained at pitch discrimination with 40 Hz amplitude modulated tones. Cerebral Cortex, 14(10), 1088-1099.

Bosshardt, H. G. (2006). Cognitive processing load as a determinant of stuttering: Summary of a research programme. Clinical linguistics & phonetics, 20(5), 371-385.

Boucher, V., Garcia, L. J., Fleurant, J., & Paradis, J. (2001). Variable efficacy of rhythm and tone in melody-based interventions: implications for the assumption of a right-hemisphere facilitation in non-fluent aphasia. Aphasiology, 15(2), 131-149.

Brady, J. P. (1971). Metronome-conditioned speech retraining for stuttering. Behavior Therapy, 2(2), 129-150.

Brandeis, D., & Lehmann, D. (1986). Event-related potentials of the brain and cognitive processes: approaches and applications. Neuropsychologia, 24(1), 151-168.

Brayton, E. R., & Conture, E. G. (1978). Effects of noise and rhythmic stimulation on the speech of stutterers. Journal of Speech and Hear- ing Research, 21, 285-294.

Bregman, A. S. (1994). Auditory scene analysis: The perceptual organization of sound. MIT press.

Bregman AS, Colantonio C, Ahad PA. (1999) Is a common grouping mechanism involved in the phenomena of illusory continuity and stream segregation? Percept Psychophys. 1999;61:195–205.

Brochard, R., D. Abecasis, D. Poter, et al. 2003. The “ticktock” of our internal clock: direct brain evidence of subjective accents in isochronous sequences. Psy- chol. Sci. 14: 362–366.

Brochard, R., Tassin, M., & Zagar, D. (2013). Got rhythm… for better and for worse. Cross-modal effects of auditory rhythm on visual word recognition. Cognition, 127(2), 214-219.

Brochard, R., Touzalin, P., Després, O., & Dufour, A. (2008). Evidence of beat perception via purely tactile stimulation. Brain research, 1223, 59-64.

Brown, P., & Marsden, C. D. (1998). What do the basal ganglia do?. The Lancet, 351(9118), 1801-1804.

Brown, S., & Weishaar, K. (2010, May). Speech is heterometric: the changing rhythms of speech. In Speech Prosody (Vol. 100074, pp. 1-4).

Brown RG. The role of cortico-striatal circuits in learning sequential information. Advances in Neurology, 80: 31–39, 1999.

Buhusi, C.V. and Meck, W.H. (2005) What makes us tick? Functional and neural mechanisms of interval timing. Nat. Rev. Neurosci. 6, 755– 765

Bull, M. (1994). Isochrony and the rhythm of conversation. In Proceedings of the Edinburgh Linguistics Department Conference (Vol. 94, pp. 5-16).

Callan, D. E., Kent, R. D., Guenther, F. H., & Vorperian, H. K. (2000). An auditory-feedback-based neural network model of speech production that is robust to developmental changes in the size and shape of the articulatory system. Journal of Speech, Language and Hearing Research, 43(3), 721.

Caruso, A. J. (1991). Neuromotor processes underlying stuttering. Speech motor control and stuttering, 101-116.

Cason, N., & Schön, D. (2012). Rhythmic priming enhances the phonological processing of speech. Neuropsychologia.

Chapin, H., Jantzen, K., Kelso, J. S., Steinberg, F., & Large, E. (2010a). Dynamic emotional and neural responses to music depend on performance expression and listener experience. PloS one, 5(12), e13812.

Chapin, H. L., Zanto, T., Jantzen, K. J., Kelso, S. J., Steinberg, F., & Large, E. W. (2010b). Neural responses to complex auditory rhythms: The role of attending. Frontiers in psychology, 1.

Chen, J. L., Zatorre, R. J., & Penhune, V. B. (2006). Interactions between auditory and dorsal premotor cortex during synchronization to musical rhythms. Neuroimage, 32(4), 1771-1781.

Chen, K., & Hasegawa-Johnson, M. (2004). How prosody improves word recognition. In Speech Prosody 2004, International Conference.

Chobert, J., Marie, C., François, C., Schön, D., & Besson, M. (2011). Enhanced passive and active processing of syllables in musician children. Journal of cognitive neuroscience, 23(12), 3874-3887.

Chomsky, N. (1956). "Three models for the description of language". IRE Transactions on Information Theory 2 (3): 113–124. doi:10.1109/TIT.1956.1056813.

Clayton, M., Sager, R., & Will, U. (2005). In time with the music: The concept of entrainment and its significance for ethnomusicology. In European meetings in ethnomusicology (Vol. 11, pp. 3-142).

Cohen, N. S., & Ford, J. (1995). The effect of musical cues on the nonpurposive speech of persons with aphasia. Journal of Music Therapy, 32, 46-46.

Colcord, R. D., & Adams, M. R. (1979). Voicing duration and vocal SPL changes associated with stuttering reduction during singing. Journal of Speech, Language and Hearing Research, 22(3), 468.

Cole, R. A., & Jakimik, J. (1980). How are syllables used to recognize words?. The Journal of the Acoustical Society of America, 67, 965.

Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. Journal of Experimental Psychology: Learning, Memory, and Cognition, 31(1), 24.

Conway, C.M., & Christiansen, M.H. (2009). Seeing and hearing in space and time: Effects of modality and presentation rate on implicit statistical learning. European Journal of Cognitive Psychology, 21, 561–580.

Conway, C. M., Karpicke, J., Anaya, E. M., Henning, S. C., Kronenberger, W. G., & Pisoni, D. B. (2011). Nonverbal cognition in deaf children following cochlear implantation: Motor sequencing disturbances mediate language delays. Developmental neuropsychology, 36(2), 237-254.

Conway, C. M., Pisoni, D. B., & Kronenberger, W. G. (2009). The importance of sound for cognitive sequencing abilities the auditory scaffolding hypothesis. Current Directions in Psychological Science, 18(5), 275-279.

Cooper, N., Cutler, A., & Wales, R. (2002). Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners. Language and speech, 45(3), 207-228.

Cooper, W. E., & Nager, R. M. (1975). Perceptuo- motor adaptation to speech: an analysis of bisyllabic utterances and a neural model. The Journal of the Acoustical Society of America, 58, 256.

Cooper GW and Meyer LB. (1960) The Rhythmic Structure of Music. Chicago: University of Chicago Press, 1960.

Coppola, V. A., & Yairi, E. (1982). Rhythmic speech training with preschool stuttering children: An experimental study. Journal of Fluency Disorders, 7(4), 447-457.

Correa, Á., & Nobre, A. C. (2008). Spatial and temporal acuity of visual perception can be enhanced selectively by attentional set. Experimental Brain Research, 189(3), 339-344.

Correa, Á., Lupiáñez, J., Madrid, E., & Tudela, P. (2006). Temporal attention enhances early visual processing: A review and new evidence from event-related potentials. Brain research, 1076(1), 116-128.

Corriveau, K., & Goswami, U. (2009). Rhythmic motor entrainment in children with speech and language impairment: Tapping to the beat. Cortex, 45, 119–130. doi:10.1016/j.cortex.2007.09.008

Corriveau K, Pasquini E, Goswami U. Basic auditory processing skills and specific language impairment: A new look at an old hypothesis. Journal of Speech, Language and Hearing Research. 2007;50:647–666.

Coull, J. T., & Nobre, A. C. (1998). Where and when to pay attention: the neural systems for directing attention to spatial locations and to time intervals as revealed by both PET and fMRI. The Journal of Neuroscience, 18(18), 7426-7435.

Coull, J. T., & Nobre, A. C. (2008). Dissociating explicit timing from temporal expectation with fMRI. Current opinion in neurobiology, 18(2), 137-144.

Coull, J. T., Davranche, K., Nazarian, B., & Vidal, F. (2012). Functional anatomy of timing differs for production versus prediction of time intervals. Neuropsychologia.

Courtney, S. M., Ungerleider, L. G., Keil, K., & Haxby, J. V. (1997). Transient and sustained activity in a distributed neural system for human working memory. Nature, 386(6625), 608-611.

Coventry, K. R., & Blackwell, T. (1994). Pragmatics in language and music. In Music Education: An Artificial Intelligence Approach (pp. 123-142). Springer London.

Cross, I. (2001). Music, cognition, culture, and evolution. Annals of the New York Academy of Science, 930, 28-42.

Cummins, F. (2002). Speech rhythm and rhythmic taxonomy. In Speech Prosody 2002, International Conference.

Cummins, F. (2003). Practice and performance in speech produced synchronously. Journal of Phonetics, 31(2), 139-148.

Cummins, F. (2006). Probing the dynamics of speech production. Methods in empirical prosody research, 3, 211.

Cummins, F. (2009a). Rhythm as an affordance for the entrainment of movement. Phonetica, 66(1-2), 15-28.

Cummins, F. (2009b). Rhythm as entrainment: The case of synchronous speech. Journal of Phonetics, 37(1), 16-28.

Cummins, F. (2011). Periodic and aperiodic synchronization in skilled action. Frontiers in human neuroscience, 5.

Cummins, F. (2012). Looking for rhythm in speech. Empirical Musicology Review 7 (1–2).

Cummins, F. (in press). Joint Speech: The Missing Link Between Speech and Music?.

Cummins, F., & Port, R. (1998). Rhythmic constraints on stress timing in English. Journal of Phonetics, 26(2), 145-171.

Cummins, F., Li, C., & Wang, B. (2013). Coupling among speakers during synchronous speaking in English and Mandarin. Journal of Phonetics, 41(6), 432-441.

Cutler, A., & Foss, D. J. (1977). On the role of sentence stress in sentence processing. Language and Speech, 20(1), 1-10.

Cutler, A., & Van Donselaar, W. (2001). Voornaam is not (really) a homophone: Lexical prosody and lexical access in Dutch. Language and speech, 44(2), 171-195.

Cutler, A., Dahan, D., & Van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. Language and speech, 40(2), 141-201.

Cutler, A., McQueen, J. M., Norris, D., & Somejuan, A. (2001). The roll of the silly ball. Language, brain and cognitive development: Essays in honor of Jacques Mehler, 181-194.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1986). The syllable's differing role in the segmentation of French and English. Journal of memory and language, 25(4), 385-400.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1992). The monolingual nature of speech segmentation by bilinguals. Cognitive psychology, 24(3), 381-410.

Cutler, A. and Butterfield, S. (1992) Rhythmic cues to speech segmentation: evidence from juncture misperception J. Mem. Lang. 31, 218–236

Dalla Bella, S., & Peretz, I. (2005). Differentiation of classical music requires little learning but rhythm. Cognition, 96(2), B65-B78.

Dalla Bella, S., Białuńska, A., & Sowiński, J. (2013). Why Movement Is Captured by Music, but Less by Speech: Role of Temporal Regularity. PloS one, 8(8), e71945.

Daltrozzo, J., & Schön, D. (2009). Conceptual processing in music as revealed by N400 effects on words and musical targets. Journal of Cognitive Neuroscience, 21(10), 1882-1892.

Darwin, C. J., & Donovan, A. (1980). Perceptual studies of speech rhythm: Isochrony and intonation. In Spoken language generation and understanding (pp. 77-85). Springer Netherlands.

Dauer, R.M. (1983) Stress-timing and syllable-timing re-analysed, Journal of Phonetics 11, pp. 51-62.

Davidow, J. H., Bothe, A. K., Andreatta, R. D., & Ye, J. (2009). Measurement of phonated intervals during four fluency-inducing conditions. Journal of Speech, Language and Hearing Research, 52(1), 188.

Delattre, P. (1938). L'accent final en français: accent d'intensité, accent de hauteur, accent de durée. The French Review, 12(2), 141-145.

Delattre, P. (1966). A comparison of syllable length conditioning among languages. ERIC Clearinghouse.

De Nil, L. F., & Kroll, R. M. (2001). Searching for the neural basis of stuttering treatment outcome: Recent neuroimaging studies. Clinical Linguistics & Phonetics, 15, 163–168.

De Nil, L. F., Kroll, R. M., Lafaille, S. J., & Houle, S. (2003). A positron emission tomography study of short- and long-term treatment effects on functional brain activation in adults who stutter. Journal of Fluency Disorders, 28, 357–380.

Deutsch, D. (1999). Grouping mechanisms in music. The psychology of music, 2, 299-348.

Dick, A. S., Solodkin, A., & Small, S. L. (2010). Neural development of networks for audiovisual speech comprehension. Brain and language, 114(2), 101-114.

Di Cristo, A. (1998). Intonation in French. Intonation systems: A survey of twenty languages, 195-218.

Dilley, L. C., & McAuley, J. D. (2008). Distal prosodic context affects word segmentation and lexical processing. Journal of Memory and Language, 59(3), 294-311.

Dilley, L. C., Mattys, S. L., & Vinke, L. (2010). Potent prosody: Comparing the effects of distal prosody, proximal prosody, and semantic context on word segmentation. Journal of Memory and Language, 63(3), 274-294.

Dimitrova, S. (1998). “Bulgarian Speech Rhythm: Stress- Timed or Syllable-Timed?”, Journal of the Inter- national Phonetic Association, vol. 27.1, 27-33, 1998

Di Pietro, M., Laganaro, M., Leemann, B., & Schnider, A. (2004). Receptive amusia: temporal auditory processing deficit in a professional musician following a left temporo-parietal lesion. Neuropsychologia, 42(7), 868-877.

Doherty, J. R., Rao, A., Mesulam, M. M., & Nobre, A. C. (2005). Synergistic effect of combined temporal and spatial expectations on visual attention. The Journal of neuroscience, 25(36), 8259-8266.

Domahs, U., Klein, E., Huber, W., & Domahs, F. (2013). Good, bad and ugly word stress–fMRI evidence for foot structure driven processing of prosodic violations. Brain and language, 125(3), 272-282.

Dominey, P. F., Hoen, M., Blanc, J. M., & Lelekov-Boissard, T. (2003). Neurological basis of language and sequential cognition: evidence from simulation, aphasia, and ERP studies. Brain and language, 86(2), 207-225.

Dowling, W. J., & Bartlett, J. C. (1981). The importance of interval information in long-term memory for melodies. Psychomusicology: Music, Mind & Brain, 1(1), 30-49.

Drake, C. (1993). Reproduction of musical rhythms by children, adult musicians, and adult nonmusicians. Perception & Psychophysics, 53(1), 25-33.

Drake, C. (1998). Psychological processes involved in the temporal organization of complex auditory sequences: universal and acquired processes. Music Percept. 16, 11–26.

Drake, C., Jones, M. R., & Baruch, C. (2000b). The development of rhythmic attending in auditory sequences: attunement, referent period, focal attending. Cognition, 77(3), 251-288.

Drake, C., Penel, A., & Bigand, E. (2000a). Tapping in time with mechanically and expressively performed music. Music Perception, 1-23.

Duncan- Johnson, C. C., & Donchin, E. (1977). On quantifying surprise: The variation of event- related potentials with subjective probability. Psychophysiology, 14(5), 456-467.

Dunham, M. J., & Newhoff, M. (1979). Melodic intonation therapy: Rewriting the song.

Dupoux, E., Pallier, C., Sebastiań, N., & Mehler, J. (1997). A destressing ‘‘deafness’’ in French? Journal of Memory and Language, 36, 406–421.

Dupoux, E., Sebastián-Gallés, N., Navarrete, E., & Peperkamp, S. (2008). Persistent stress ‘deafness’: The case of French learners of Spanish. Cognition, 106(2), 682-706.

Eckstein, K., & Friederici, A. D. (2005). Late interaction of syntactic and prosodic processes in sentence comprehension as revealed by ERPs. Cognitive Brain Research, 25(1), 130-143.

Eckstein, K., & Friederici, A. D. (2006). It's early: event-related potential evidence for initial interaction of syntax and prosody in speech comprehension. Journal of Cognitive Neuroscience, 18(10), 1696-1711.

Ellis, R. J., & Jones, M. R. (2010). Rhythmic context modulates foreperiod effects. Attention, Perception, & Psychophysics, 72(8), 2274-2288.

Elman, J. L. (1981). Effects of frequency- shifted feedback on the pitch of vocal productions. The Journal of the Acoustical Society of America, 70, 45.

Elsinger CL, Rao SM, Zimbelman JL, Reynolds NC, Blindauer KA, Hoffman RG: Neural basis for impaired time reproduction in Parkinson’s disease: An fMRI study. J Int Neuropsychol Soc 2003, 9:1088-1098.

Escoffier, N., Sheng, D. Y. J., & Schirmer, A. (2010). Unattended musical beats enhance visual processing. Acta psychologica, 135(1), 12-16.

Essens, P. J., & Povel, D. J. (1985). Metrical and nonmetrical representations of temporal patterns. Perception & Psychophysics, 37(1), 1-7.

Fadiga, L., Craighero, L., & D’Ausilio, A. (2009). Broca's area in language, action, and music. Annals of the New York Academy of Sciences, 1169(1), 448-458.

Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: a TMS study. European Journal of Neuroscience, 15(2), 399-402.

Farrugia, N., Benoit, C. E., Harding, E., Kotz, S. A., & Dalla Bella, S. (2012). BAASTA: Battery for the Assessment of Auditory Sensorimotor and Timing Abilities. Institute for Human Cognitive and Brain Sciences (n, 18, 9.

Fedorenko, E., Patel, A., Casasanto, D., Winawer, J., & Gibson, E. (2009). Structural integration in language and music: Evidence for a shared system. Memory & cognition, 37(1), 1-9.

Fowler, C. A., Smith, M. R., & Tassinary, L. G. (1986). Perception of syllable timing by prebabbling infants. Journal of the Acoustical Society of America, 79, 814–825.

Fraisse P. Rhythm and tempo. In: Deutsch D, editor. The Psychology of Music. Academic Press; London: 1982. pp. 149–180.

Francois, C., & Schön, D. (2010). Learning of musical and linguistic structures: comparing event-related potentials and behavior. Neuroreport, 21(14), 928-932.

Francois, C., & Schön, D. (2011). Musical expertise boosts implicit learning of both musical and linguistic structures. Cerebral Cortex, 21(10), 2357-2365.

Franklin, M. S., Moore, K. S., Yip, C. Y., Jonides, J., Rattray, K., & Moher, J. (2008). The effects of musical training on verbal memory. Psychology of Music, 36(3), 353-365.

Fransella, F., & Beech, H. R. (1965). An experimental analysis oe the effect of rhythm on the speech of stutterers. Behaviour Research and Therapy, 3(3), 195-201.

François, C., Chobert, J., Besson, M., & Schön, D. (2012). Music training for the development of speech segmentation. Cerebral Cortex.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. Trends in cognitive sciences, 6(2), 78-84.

Friederici, A. D., & Kotz, S. A. (2003). The brain basis of syntactic processes: functional imaging and lesion studies. Neuroimage, 20, S8-S17.

Fries, W., & Swihart, A. A. (1990). Disturbance of rhythm sense following right hemisphere damage. Neuropsychologia, 28(12), 1317-1323.

Fujioka, T., Ross, B., Kakigi, R., Pantev, C., & Trainor, L. J. (2006). One year of musical training affects development of auditory cortical-evoked fields in young children. Brain, 129(10), 2593-2608.

Fujioka T., Trainor L. J., Large E. W., Ross B. (2009). Beta and gamma rhythms in human auditory cortex during musical beat processing. Ann. N. Y. Acad. Sci. 1169, 89–92. doi: 10.1111/j.1749-6632.2009.04779.x.

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. Psychonomic bulletin & review, 13(3), 361-377.

Gaser, C., & Schlaug, G. (2003). Brain structures differ between musicians and non-musicians. The Journal of Neuroscience, 23(27), 9240-9245.

Geiser, E., Sandmann, P., Jäncke, L., & Meyer, M. (2010). Refinement of metre perception–training increases hierarchical metre processing. European Journal of Neuroscience, 32(11), 1979-1985.

Geiser, E., Ziegler, E., Jancke, L., & Meyer, M. (2009). Early electrophysiological correlates of meter and rhythm processing in music perception. Cortex, 45(1), 93-102.

Gentilucci M, Dalla Volta R. Spoken language and arm gestures are controlled by the same motor control system. Q. J. Exp. Psychol. (Colchester) 2008;61:944–957.

Georgiewa, P., Rzanny, R., Gaser, C., Gerhard, U. J., Vieweg, U., Freesmeyer, D., Mentzel, H, Kaiser, W. A. & Blanz, B. (2002). Phonological processing in dyslexic children: a study combining functional imaging and event related potentials. Neuroscience letters, 318(1), 5-8.

Gibbon, D., & Gut, U. (2001). Measuring speech rhythm. In INTERSPEECH (pp. 95-98).

Gill, S., Thompson, M. R., & Himberg, T. Body Rhythmic Entrainment and Pragmatics in Musical and Linguistic Improvisation Tasks. Proceedings on the 12th International Conference on Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music, July 23 - 28, 2012, Thessaloniki, Greece. Cambouropoulos, E., Tsougras, C., Mavromatis, P. & Pastiadis, K (Editors).

Gill, S. P. (2012). Rhythmic synchrony and mediated interaction: Towards a framework of rhythm in embodied interaction. AI & society, 27(1), 111-127.

Giraud, A. L., & Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. Nature neuroscience, 15(4), 511-517.

Giraud, A. L., Neumann, K., Bachoud-Levi, A. C., von Gudenberg, A. W., Euler, H. A., Lanfermann, H., & Preibisch, C. (2008). Severity of dysfluency correlates with basal ganglia activity in persistent developmental stuttering. Brain and language, 104(2), 190-199.

Goldinger, S. D., & Azuma, T. (2003). Puzzle-solving science: The quixotic quest for units in speech perception. Journal of Phonetics, 31(3), 305-320.

Gordon, R. L., Magne, C. L., & Large, E. W. (2011). EEG Correlates of Song Prosody: a new look at the relationship between linguistic and musical rhythm. Frontiers in psychology, 2.

Gow Jr, D. W., & Gordon, P. C. (1993). Coming to terms with stress: Effects of stress location in sentence processing. Journal of psycholinguistic research, 22(6), 545-578.

Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. Papers in laboratory phonology, 7(515-546).

Graeber S, Hertrich I, Daum I, Spieker S, Ackermann H: Speech perception deficits in Parkinson’s disease: underestimation of time intervals compromises identification of durational phonetic contrasts. Brain Lang 2002, 82:65-74.

Grahn, J. A. (2009). The role of the basal ganglia in beat perception. Annals of the New York Academy of Sciences, 1169(1), 35-45.

Grahn, J. A. (2012a). Neural mechanisms of rhythm perception: Current findings and future perspectives. Topics in cognitive science, 4(4), 585-606.

Grahn, J. A. (2012b). See what I hear? Beat perception in auditory and visual rhythms. Experimental brain research, 220(1), 51-61.

Grahn, J. A., & Brett, M. (2007). Rhythm and beat perception in motor areas of the brain. Journal of Cognitive Neuroscience, 19(5), 893-906.

Grahn, J. A., & Brett, M. (2009). Impairment of beat-based rhythm discrimination in Parkinson's disease. Cortex, 45(1), 54-61.

Grahn, J. A., & Brett, M. (2009). Impairment of beat-based rhythm discrimination in Parkinson’s disease. Cortex, 45, 54–61. doi:10.1016/ j.cortex.2008.01.005

Grahn, J. A., & McAuley, J. D. (2009). Neural bases of individual differences in beat perception. Neuroimage, 47(4), 1894-1903.

Grahn, J. A., & Rowe, J. B. (2009). Feeling the beat: premotor and striatal interactions in musicians and nonmusicians during beat perception. The Journal of Neuroscience, 29(23), 7540-7548.

Grahn, J. A., & Rowe, J. B. (2013). Finding and feeling the musical beat: Striatal dissociations between detection and prediction of regularity. Cerebral cortex, 23(4), 913-921.

Grahn, J. A., Henry, M. J., & McAuley, J. D. (2011). FMRI investigation of cross-modal interactions in beat perception: Audition primes vision, but not vice versa. NeuroImage, 54(2), 1231-1243.

Graybiel AM. (1997) The basal ganglia and cognitive pattern generators.Schizophrenia Bulletin, 23: 459–469, 1997.

Greenberg, S., Carvey, H., Hitchcock, L., & Chang, S. (2003). Temporal properties of spontaneous speech—a syllable-centric perspective. Journal of Phonetics, 31(3), 465-485.

Greenberg, S., Marsh, J. T., Brown, W. S. & Smith, J. C. Neural temporal coding of low pitch. I. Human frequency-following responses to complex tones. Hear Res 25, 91-114 (1987).

Griffin, I. C., Miniussi, C., & Nobre, A. (2002). Multiple mechanisms of selective attention: differential modulation of stimulus processing by attention to space or time. Neuropsychologia, 40, 2325–2340.

Griffiths, T. D., & Warren, J. D. (2002). The planum temporale as a computational hub. Trends in neurosciences, 25(7), 348-353.

Grosjean, F., & Gee, J. P. (1987). Prosodic structure and spoken word recognition. Cognition, 25(1), 135-155.

Guberina, P., & Asp, C. W. (1981). The verbo-tonal method for rehabilitating people with communication problems. Monograph 13, World Rehabilitation.

Guberina, P. E. T. A. R. (1963, June). Verbotonal method and its application to the rehabilitation of the deaf. In International congress on education of the deaf. Washington, DC: US government printing office (pp. 279-293).

Guenther, F. H. (2006). Cortical interactions underlying the production of speech sounds. Journal of communication disorders, 39(5), 350-365.

Hagoort, P. (2006). On Broca, brain, and binding. Broca's Region. Oxford Univ. Press, Oxford, 242-253.

Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. Language and cognitive processes, 8(4), 439-483.

Hahn, L. D. (1999). Native speakers' reactions to non-native stress in English discourse (Doctoral dissertation, University of Illinois at Urbana-Champaign).

Hannon, E.E. & S.E. Trehub. 2005. Metrical categories in infancy and adulthood. Psychol. Sci. 16: 48–55.

Hannon, E. E. (2009). Perceiving speech rhythm in music: Listeners classify instrumental songs according to language of origin. Cognition, 111(3), 403-409.

Harrington DL, Haaland KY, Hermanowicz N: Temporal processing in the basal ganglia. Neuropsychology 1998, 12:3-12.

Harrington DL, Lee RR, Boyd LA, Rapcsak SZ, Knight RT: Does the representation of time depend on the cerebellum? Effect of cerebellar stroke. Brain 2004, 127:561-574.

Hay, J. S., & Diehl, R. L. (2007). Perception of rhythmic grouping: Testing the iambic/trochaic law. Perception & Psychophysics, 69(1), 113-122.

Healey, E. C., Mallard III, A. R., & Adams, M. R. (1976). Factors contributing to the reduction of stuttering during singing. Journal of Speech, Language and Hearing Research, 19(3), 475.

Herdener, M., Humbel, T., Esposito, F., Habermeyer, B., Cattapan-Ludewig, K., & Seifritz, E. (2012). Jazz Drummers Recruit Language-Specific Areas for the Processing of Rhythmic Structure. Cerebral Cortex.

Herholz, S. C., & Zatorre, R. J. (2012). Musical training as a framework for brain plasticity: behavior, function, and structure. Neuron, 76(3), 486-502.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. Nature Reviews Neuroscience, 8(5), 393-402.

Hickok, G., Okada, K., & Serences, J. T. (2009). Area Spt in the human planum temporale supports sensory-motor integration for speech processing. Journal of Neurophysiology, 101(5), 2725-2732.

Hickok, G. and Poeppel, D. (2000) Towards a functional neuroanatomy of speech perception. Trends Cogn. Sci. 4, 131–138

Hirsh, I. J., Monahan, C. B., Grant, K. W., & Singh, P. G. (1990). Studies in auditory timing: 1. Simple patterns. Perception & psychophysics, 47(3), 215-226.

Hirsh-Pasek, K., Kemler Nelson, D. G., Jusczyk, P. W., Cassidy, K. W., Druss, B., & Kennedy, L. (1987). Clauses are perceptual units for young infants. Cognition, 26(3), 269-286.

Ho, Y. C., Cheung, M. C., & Chan, A. S. (2003). Music training improves verbal but not visual memory: cross-sectional and longitudinal explorations in children. Neuropsychology, 17(3), 439.

Honing, H. (2012) Without it no music: beat induction as a fundamental musical trait. Annals of the New York Academy of Sciences 1252:1, 85-91

Honing, H., & Ladinig, O. (2009). Exposure influences expressive timing judgments in music. Journal of Experimental Psychology: Human Perception and Performance, 35(1), 281.

Honing, H., O. Ladinig, I. Winkler & G. Ha ́den. (2009). Is beat induction innate or learned? Probing emergent meter perception in adults and newborns using event-related brain potentials (ERP). Ann. N.Y. Acad. Sci. 1169: 93–96.

Honing, H. 2011. The Illiterate Listener. On Music Cognition, Musicality and Methodology. Amsterdam University Press.Amsterdam.

Hopyan, T., Schellenberg, E. G., & Dennis, M. (2009). Perception of strong-meter and weak-meter rhythms in children with spina bifida meningomyelocele. Journal of the International Neuropsychological Society, 15(4), 521.

Horn, D. L., Pisoni, D. B., & Miyamoto, R. T. (2006). Divergence of fine and gross motor skills in prelingually deaf children: implications for cochlear implantation. The Laryngoscope, 116(8), 1500-1506.

Howie, P. M. (1981). Concordance for stuttering in monozygotic and dizygotic twin pairs. Journal of Speech, Language and Hearing Research, 24(3), 317.

Huang, C. and Liu, G. (1990) Organization of the auditory area in the posterior cerebellar vermis of the cat. Exp. Brain Res. 81, 377–383

Hubbard, C. P. (1998). Stuttering, stressed syllables, and word onsets. Journal of Speech, Language and Hearing Research, 41(4), 802.

Huffman, R. F., & Henson Jr, O. W. (1990). The descending auditory pathway and acousticomotor systems: connections with the inferior colliculus. Brain research. Brain research reviews, 15(3), 295.

Hulse, S. H., & Page, S. C. (1988). Toward a comparative psychology of music perception. Music Perception, 5, 427-452.

Huron, D. (2001). Is music an evolutionary adaptation? Annals of the New York Academy of Science, 930, 43-61.

Huron, D., & Ollen, J. (2003). Agogic contrast in French and English themes: Further support for Patel and Daniele (2003). Music Perception: An Interdisciplinary Journal, 21(2), 267-271.

Huss, M., Verney, J. P., Fosker, T., Mead, N., & Goswami, U. (2011). Music, rhythm, rise time perception and developmental dyslexia: perception of musical meter predicts reading and phonology. Cortex, 47(6), 674-689.

Ibbotson, N. R., & Morton, J. (1981). Rhythm and dominance. Cognition, 9(2), 125-138.

Ingham, R. J. (1984). Stuttering and behavior therapy: Current status and experimental foundations. San Diego, CA: College-Hill Press.

Iversen, J. R., Patel, A. D., & Ohgushi, K. (2008). Perception of rhythmic grouping depends on auditory experience. The Journal of the Acoustical Society of America, 124, 2263.

Iversen, J. R., Repp, B. H., & Patel, A. D. (2009). Top- Down Control of Rhythm Perception Modulates Early Auditory Responses. Annals of the New York Academy of Sciences, 1169(1), 58-73.

Ivry, R. B., & Spencer, R. (2004). The neural representation of time. Current opinion in neurobiology, 14(2), 225-232.

Ivry RB, Keele SW: Timing functions of the cerebellum. J Cogn Neurosci 1989, 1:136-152.

James, C. E., Oechslin, M. S., Van De Ville, D., Hauert, C. A., Descloux, C., & Lazeyras, F. (2013). Musical training intensity yields opposite effects on grey matter density in cognitive versus sensorimotor networks. Brain Structure and Function, 1-14.

Jancke, L., Shah, N.J., Peters, M., 2000. Cortical activations in primary and secondary motor areas for complex bimanual movements in professional pianists. Cogn. Brain Res. 10, 177–183.

Jeffries, K. J., Fritz, J. B., & Braun, A. R. (2003). Words in melody: an H215O PET study of brain activation during singing and speaking. Neuroreport, 14(5), 749.

Jensen O., Kaiser J., Lachaux J. P. (2007). Human gamma-frequency oscillations associated with attention and memory. Trends Neurosci. 30, 317–324. doi: 10.1016/j.tins.2007.05.001.

Jentschke, S., Koelsch, S., Sallat, S., & Friederici, A. D. (2008). Children with specific language impairment also show impairment of music-syntactic processing. Journal of Cognitive Neuroscience, 20(11), 1940-1951.

Johnson, K., Nicol, T. & Kraus, N. The Brainstem Response to Speech: A Biological Marker of Auditory Processing. Ear Hear 26, 424-34 (2005).

Jones, M. R. (1993). Dynamics of musical patterns: How do melody and rhythm fit together?. In Based on a talk presented at the meeting of the American Association for the Advancement of Science in May 1986.. Lawrence Erlbaum Associates, Inc.

Jones, M. R. (2009). Musical time. In S. Hallam, I. Cross, & M. Thaut, (Eds.), Oxford handbook of music psychology (pp. 81-92). Oxford: Oxford University Press.

Jones, M. R., & Boltz, M. (1989). Dynamic attending and responses to time. Psychological review, 96(3), 459.

Jones, M. R., Boltz, M., & Kidd, G. (1982). Controlled attending as a function of melodic and temporal context. Perception & Psychophysics, 32(3), 211-218.

Jones, M. R., Moynihan, H., MacKenzie, N., & Puente, J. (2002). Temporal aspects of stimulus-driven attending in dynamic arrays. Psychological Science, 13(4), 313-319.

Jones MR (1993) Dynamics of musical patterns: How do melody and rhythm fit together? In: TJ TigheWJ Dowling. Psychology and music: The understanding of melody and rhythm. Hillsdale, NJ: Erlbaum. pp. 67-92.

Jongsma, M. L., Desain, P., & Honing, H. (2004). Rhythmic context influences the auditory evoked potentials of musicians and nonmusicians. Biological Psychology, 66(2), 129-152.

Jongsma, M. L., Eichele, T., Quiroga, R. Q., Jenks, K. M., Desain, P., Honing, H., & Van Rijn, C. M. (2005). Expectancy effects on omission evoked potentials in musicians and non-musicians. Psychophysiology, 42(2), 191-201.

Jongsma, M. L., Quiroga, R. Q., & van Rijn, C. M. (2004). Rhythmic training decreases latency-jitter of omission evoked potentials (OEPs) in humans. Neuroscience letters, 355(3), 189-192.

Jungers, M. K., Palmer, C., & Speer, S. R. (2002). Time after time: The coordinating influence of tempo in music and speech. Cognitive Processing, 1(2), 21-35.

Jusczyk, P. W., Cutler, A., & Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of English words. Child development, 64(3), 675-687.

Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. Cognitive psychology, 39(3), 159-207.

Juslin, P. N., & Sloboda, J. A. (2001). Music and emotion: Theory and research. Oxford University Press.

Jäncke, L., Loose, R., Lutz, K., Specht, K., & Shah, N. J. (2000). Cortical activations during paced finger-tapping applying visual and auditory pacing stimuli. Cognitive Brain Research, 10(1), 51-66.

Jääskeläinen, I. P., Ahveninen, J., Belliveau, J. W., Raij, T., & Sams, M. (2007). Short-term plasticity in auditory cognition. Trends in neurosciences, 30(12), 653-661.

Jääskeläinen, I. P., Ahveninen, J., Bonmassar, G., Dale, A. M., Ilmoniemi,R.J.,&Leva □nen,S.,etal.(2004).Humanposteriorau- ditory cortex gates novel sounds to consciousness. Proceedings of the National Academy of Sciences of the USA, 17, 6809–6814.

Karniol, R. (1995). Stuttering, language, and cognition: A review and a model of stuttering as suprasegmental sentence plan alignment (SPA). Psychological Bulletin, 117(1), 104.

Keenan, J. P., Thangaraj, V., Halpern, A. R., & Schlaug, G. (2001). Absolute pitch and planum temporale. Neuroimage, 14(6), 1402-1408.

Kell, C. A., Neumann, K., von Kriegstein, K., Posenenske, C., von Gudenberg, A. W., Euler, H., & Giraud, A. L. (2009). How the brain repairs stuttering. Brain, 132(10), 2747-2760.

Kell, C. A., Neumann, K., von Kriegstein, K., Posenenske, C., von Gudenberg, A. W., Euler, H., & Giraud, A. L. (2009). How the brain repairs stuttering. Brain, 132(10), 2747-2760.

Keller, P. (1999). Attending in complex musical interactions: The adaptive dual role of meter. Australian Journal of Psychology, 51, 166-175.

Kendon, A. (2004). Gesture: Visible action as utterance. Cambridge University Press.

Kester, D. B., Saykin, A. J., Sperling, M. R., O'Connor, M. J., Robinson, L. J., & Gur, R. C. (1991). Acute effect of anterior temporal lobectomy on musical processing. Neuropsychologia, 29(7), 703-708.

Kidd, K. (1984). Stuttering as a genetic disorder. Nature and treatment of stuttering: New directions, 149-169.

Kluender, K.R. and Greenberg, S. (1989) A specialization for speech perception? Science 244, 1530

Kochanski, G., & Orphanidou, C. (2008). What marks the beat of speech?. The Journal of the Acoustical Society of America, 123, 2780.

Koelsch, S. (2005). Neural substrates of processing syntax and semantics in music. Current opinion in neurobiology, 15(2), 207-212.

Koelsch, S., Gunter, T. C., Wittfoth, M., & Sammler, D. (2005). Interaction between syntax processing in language and in music: an ERP study. Journal of Cognitive Neuroscience, 17(10), 1565-1577.

Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., & Friederici, A. D. (2004). Music, language and meaning: brain signatures of semantic processing. Nature neuroscience, 7(3), 302-307.

Koelsch S, Kasper E, Sammler D, Schulze K, Gunter T, Friederici AD. Music, language and meaning: brain signatures of semantic processing. Nat Neurosci. 2004;16:479-495.

Koelsch S., Gunter T.C., von Cramon D.Y., Zysset S., Lohmann G., & Friederici A.D. (2002). Bach speaks: A cortical 'language-network' serves the processing of music. Neuroimage, 17:956-966.

Koeneke, S., Lutz, K., Wüstenberg, T., & Jäncke, L. (2004). Long-term training affects cerebellar processing in skilled keyboard players. Neuroreport, 15(8), 1279-1282.

Kotz, S. A., & Schwartze, M. (2010). Cortical speech processing unplugged: a timely subcortico-cortical framework. Trends in cognitive sciences, 14(9), 392-399.

Kotz, S. A., & Schwartze, M. (2011). Differential input of the supplementary motor area to a dedicated temporal processing network: functional and clinical implications. Frontiers in integrative neuroscience, 5.

Kotz, S. A., Frisch, S., von Cramon, D. Y., & Friederici, A. D. (2003). Syntactic language processing: ERP lesion data on the role of the basal ganglia. Journal of the International Neuropsychological Society, 9(7), 1053-1060.

Kotz, S. A., Gunter, T. C., & Wonneberger, S. (2005). The basal ganglia are receptive to rhythmic compensation during auditory syntactic processing: ERP patient data. Brain and Language, 95(1), 70-71.

Kotz, S.A. Schwartze & Schmidt kassow (2009) Non-motor basal ganglia functions: A review and proposal for a model of sensory predictability in auditory language perception. Cortex 45, 982–990

Krahmer, E., & Swerts, M. (2007). The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. Journal of Memory and Language, 57(3), 396-414.

Kral, A., & Eggermont, J. J. (2007). What's to lose and what's to learn: development under auditory deprivation, cochlear implants and limits of cortical plasticity. Brain Research Reviews, 56(1), 259-269.

Kral, A., Hartmann, R., Tillein, J., Heid, S., Klinke, R., (2000). Congenital auditory deprivation reduces synaptic activity within the auditory cortex in a layer-specific manner. Cereb. Cortex 10, 714–726

Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. Nature Reviews Neuroscience, 11(8), 599-605.

Krumhansl, C. L. (2000). Rhythm and pitch in music cognition. Psychological bulletin, 126(1), 159.

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. Nature reviews neuroscience, 5(11), 831-843.

Kung, S. J., Tzeng, O. J., Hung, D. L., & Wu, D. H. (2011). Dynamic allocation of attention to metrical and grouping accents in rhythmic sequences. Experimental Brain Research, 210(2), 269-282.

Kusumoto, K., & Moreton, E. (1997). Native language determines the parsing of nonlinguistic rhythmic stimuli. The Journal of the Acoustical Society of America, 102, 3204.

Kühnis, J., Elmer, S., Meyer, M., & Jäncke, L. (2013). The encoding of vowels and temporal speech cues in the auditory cortex of professional musicians: An EEG study. Neuropsychologia.

Ladefoged, P. (1975) A Course in Phonetics. New York: Harcourt Brace Jovanovich.

Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., & Schroeder, C. E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. science, 320(5872), 110-113.

Lancker, D. V., & Sidtis, J. J. (1992). The identification of affective-prosodic stimuli by left- and right-hemisphere-damaged subjects: all errors are not created equal. Journal of Speech, Language and Hearing Research, 35(5), 963.

Langacker, R. W. (1987). Foundations of Cognitive Grammar: theoretical prerequisites. Volume 1 (Vol. 1). Stanford university press.

Lange, K. (2009). Brain correlates of early auditory processing are attenuated by expectations for time and pitch. Brain and cognition, 69(1), 127-137.

Large, E. W. (2008). Resonating to musical rhythm: theory and experiment. Psychology of time, 189-232.

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. Psychological review, 106(1), 119.

Large, E. W., & Kolen, J. F. (1994). Resonance and the perception of musical meter. Connection science, 6(2-3), 177-208.

Large, E. W., & Palmer, C. (2002). Perceiving temporal regularity in music. Cognitive Science, 26(1), 1-37.

Larrouy-Maestri, P., Lévêque, Y., Schön, D., Giovanni, A., & Morsomme, D. (2012). The Evaluation of Singing Voice Accuracy: A Comparison Between Subjective and Objective Methods. Journal of Voice.

Lehiste, I. (1977). Isochrony reconsidered. Journal of phonetics.

Leonard, T., & Cummins, F. (2011). The temporal relation between beat gestures and speech. Language and Cognitive Processes, 26(10), 1457-1471.

Leong, V., & Goswami, U. (2013). Assessment of Rhythmic Entrainment at Multiple Timescales in Dyslexia: Evidence for Disruption to Syllable Timing. Hearing research.

LERDAHL, F. A., & Jackendoff, R. S. (1983). A generative theory of tonal music. The MIT Press.

Levitin, D. J. & Menon, V. (2003). Musical structure is processed in “language” areas of the brain: a possible role for Brodmann Area 47 in temporal coherence. Neuroimage 20, 2142-52.

Lewis, P. A., Wing, A. M., Pope, P. A., Praamstra, P., & Miall, R. C. (2004). Brain activity correlates differentially with increasing temporal complexity of rhythms during initialisation, synchronisation, and continuation phases of paced finger tapping. Neuropsychologia, 42(10), 1301-1312.

Leyden, K. V., & Heuven, V. V. (1996). Lexical stress and spoken word recognition: Dutch vs. English.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. Cognition, 21(1), 1-36.

Liberman, A. M., & Whalen, D. H. (2000). On the relation of speech to language. Trends in cognitive sciences, 4(5), 187-196.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. Psychological review, 74(6), 431.

Lidji, P., Palmer, C., Peretz, I., & Morningstar, M. (2011b). Listeners feel the beat: Entrainment to English and French speech rhythms. Psychonomic bulletin & review, 18(6), 1035-1041.

Lidji, P., Palmer, C., Peretz, I., & Morningstar, M. (2011a). Entrainment to speech and song..International Symposium on Performance ScienceISBN 978-94-90306-02-1 www.performingscience.org

Lieberman, Philip. (2000) Human Language and the Reptilian Brain: The Subcortical Bases of Speech, Syntax, and Thought. Cambridge, Mass.: Harvard University Press.

Liegeois-Chauvel, C., Musolino, A., Badier, J. M., Marquis, P., & Chauvel, P. (1994). Evoked potentials recorded from the auditory cortex in man: evaluation and topography of the middle latency components. Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section, 92(3), 204-214.

Liégeois-Chauvel, C., Peretz, I., Babaï, M., Laguitton, V., & Chauvel, P. (1998). Contribution of different cortical areas in the temporal lobes to music processing. Brain, 121(10), 1853-1867.

London, J. (2012). Hearing in time. Oxford University Press.

London, J., 2004. Hearing in Time: Psychological Aspects of Musical Meter. Oxford University Press, USA.

Lotze, M., Scheler, G., Tan, H. R., Braun, C., & Birbaumer, N. (2003). The musician's brain: functional imaging of amateurs and professionals during performance and imagery. Neuroimage, 20(3), 1817-1829.

Loui, P., Guenther, F. H., Mathys, C., & Schlaug, G. (2008). Action–perception mismatch in tone-deafness. Current Biology, 18(8), R331-R332.

Lu, C., Peng, D., Chen, C., Ning, N., Ding, G., Li, K., ... & Lin, C. (2010). Altered effective connectivity and anomalous anatomy in the basal ganglia-thalamocortical circuit of stuttering speakers. Cortex, 46(1), 49-67.

Ludlow, C. L., & Loucks, T. (2004). Stuttering: a dynamic motor control disorder. Journal of fluency disorders, 28(4), 273-295.

Ludlow, C. L., Connor, N. P., & Bassich, C. J. (1987). Speech timing in Parkinson's and Huntington's disease. Brain and language, 32(2), 195-214.

Lukatela, G., & Turvey, M. T. (1996). Inhibition of naming by rhyming primes. Perception & Psychophysics, 58(6), 823-835.

Madison, G., & Merker, B. (2002). On the limits of anisochrony in pulse attribution. Psychological Research, 66(3), 201-207.

Maegele, M., Lippert- Gruener, M., Ester- Bode, T., Garbe, J., Bouillon, B., Neugebauer, E., ... & Angelov, D. N. (2005). Multimodal early onset stimulation combined with enriched environment is associated with reduced CNS lesion volume and enhanced reversal of neuromotor dysfunction after traumatic brain injury in rats. European Journal of Neuroscience, 21(9), 2406-2418.

Maess, B., Koelsch, S., Gunter, T., & Friederici, A.D. (2001). Musical syntax is processed in Broca’s area: an MEG study. Nature Neuroscience, 4:540-545.

Magne, C., Astésano, C., Aramaki, M., Ystad, S., Kronland-Martinet, R., & Besson, M. (2007). Influence of syllabic lengthening on semantic processing in spoken French: behavioral and electrophysiological evidence. Cerebral cortex, 17(11), 2659-2668.

Magne, C., Schön, D., & Besson, M. (2006). Musician children detect pitch violations in both music and language better than nonmusician children: behavioral and electrophysiological approaches. Journal of Cognitive Neuroscience, 18(2), 199-211.

Mairano, P., & Romano, A. (2008). A comparison of four rhythm metrics for six languages. In Poster presented at the conference on Empirical Approaches to Speech Rhythm. University College London.

Marie, C., Magne, C., & Besson, M. (2011). Musicians and the metric structure of words. Journal of Cognitive Neuroscience, 23(2), 294-305.

Marques, C., Moreno, S., & Besson, M. (2007). Musicians detect pitch violation in a foreign language better than nonmusicians: Behavioral and electrophysiological evidence. Journal of Cognitive Neuroscience, 19(9), 1453-1463.

Marschark, M. (2006). Intellectual functioning of deaf adults and children: Answers and questions. European Journal of Cognitive Psychology, 18(1), 70-89.

Max, L., & Yudman, E. M. (2003). Accuracy and variability of isochronous rhythmic timing across motor systems in stuttering versus nonstuttering individuals. Journal of Speech, Language and Hearing Research, 46(1), 146.

Max, L., Guenther, F. H., Gracco, V. L., Ghosh, S., S., & Wallace, M. E. (2004). Unstable or Insufficiently Activated Internal Models and Feedback--- Biased Motor Control as Sources of Dysfluency: A Theoretical M. Contemporary Issues in Communication Science and Disorders, 31, 105–122.

McClelland, J. L., & Intelligence, A. (1987). The case for interactionism in language processing.

McDermott, J., & Hauser, M. (2005). The origins of music: Innateness, uniqueness, and evolution. Music Perception: An interdisciplinary journal, 23(1), 29-59.

McIntosh, G. C., Brown, S. H., Rice, R. R., & Thaut, M. H. (1997). Rhythmic auditory-motor facilitation of gait patterns in patients with Parkinson's disease. Journal of Neurology, Neurosurgery & Psychiatry, 62(1), 22-26.

McNeill, D. (1992). Hand and mind: What gestures reveal about thought. University of Chicago Press.

Meck WH, Benson AM: Dissecting the brain’s internal clock: □ how frontal-striatal circuitry keeps time and shifts attention. Brain Cogn 2002, 48:195-211.

Mehler, J., & Christophe, A. (1995). Maturation and learning of language in the first year of life, in Gazzaniga, M.S. (ed.), The Cognitive Neurosciences; a handbook for the field, Cambridge (MA), Mit Press, 943-54

Meister, I. G., Krings, T., Foltys, H., Boroojerdi, B., Müller, M., Töpper, R., & Thron, A. (2004). Playing piano in the mind—an fMRI study on music imagery and performance in pianists. Cognitive Brain Research, 19(3), 219-228.

Menning H, Roberts LE, Pantev C. Plastic changes in the auditory cortex induced by intensive frequency discrimination training. Neuroreport. 2000;11:817-822.

Merlo, S., and L. L. Mansur. 2004. Descriptive discourse: topic familiarity and disfluencies. Journal of Communication Disorders 37.489-503.

Meyer, L., 1956. Emotion and Meaning in Music. University of Chicago Press, Chicago.

Meyer, M., Elmer, S., Ringli, M., Oechslin, M. S., Baumann, S., & Jancke, L. (2011). Long-term exposure to music enhances the sensitivity of the auditory system in children. European Journal of Neuroscience, 34(5), 755-765.

Miller, J.E, Carlson, L. A., & McAuley, J.D. (2013). When what you hear influences when you see: Listening to an auditory rhythm influences the temporal allocation of visual attention. Psychological Science, 24, 11-18.

Milovanov, R., Huotilainen, M., Esquef, P. A., Alku, P., Välimäki, V., & Tervaniemi, M. (2009). The role of musical aptitude and language skills in preattentive duration processing in school-aged children. Neuroscience letters, 460(2), 161-165.

Moreno, S., & Besson, M. (2006). Musical training and language- related brain electrical activity in children. Psychophysiology, 43(3), 287-291.

Moreno, S., Marques, C., Santos, A., Santos, M., & Besson, M. (2009). Musical training influences linguistic abilities in 8-year-old children: more evidence for brain plasticity. Cerebral Cortex, 19(3), 712-723.

Morgan, J. L., & Saffran, J. R. (1995). Emerging integration of sequential and suprasegmental information in preverbal speech segmentation. Child development, 66(4), 911-936.

Morton, J., Marcus, S., & Frankish, C. (1976). Perceptual centers (P-centers). Psychological Review, 83(5), 405.

Muneaux, M., Ziegler, J. C., Truc, C., Thomson, J., & Goswami, U. (2004). Deficits in beat perception and dyslexia: Evidence from French. Neu- roReport, 15, 1255–1259. doi:10.1097/01.wnr.0000127459.31232.c4

Murayama, J., Kashiwagi, T., Kashiwagi, A., & Mimura, M. (2004). Impaired pitch production and preserved rhythm production in a right brain-damaged patient with amusia. Brain and Cognition, 56(1), 36-42.

Musacchia, G., Sams, M., Skoe, E., & Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. Proceedings of the National Academy of Sciences, 104(40), 15894-15898.

Musacchia, G., Strait, D., & Kraus, N. (2008). Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians. Hearing research, 241(1), 34-42.

Månsson, H. (2000). Childhood stuttering. Journal of Fluency Disorders, 25(1), 47– 57.

Nathan, G. S. (2011, October). Phonemes as mental categories. In Proceedings of the Annual Meeting of the Berkeley Linguistics Society (Vol. 12).

Nazzi, T., & Ramus, F. (2003). Perception and acquisition of linguistic rhythm by infants. Speech Communication, 41(1), 233-243.

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: toward an understanding of the role of rhythm. Journal of Experimental Psychology: Human perception and performance, 24(3), 756.

Nearey, T. M. (1997). Speech perception as pattern recognition. Journal of the Acoustical Society of America, 101, 3241–3254.

Nelson, D. G. K., Hirsh-Pasek, K., Jusczyk, P. W., & Cassidy, K. W. (1989). How the prosodic cues in motherese might assist language learning. Journal of child Language, 16(01), 55-68.

Nespor, M. (1990). On the rhythm parameter in phonology. In I. M. Roca (Ed.), Logical issues in language acquisition (pp. 157-175). Dordrecht: Foris.

Neumann et al., 2003 K. Neumann, H.A. Euler, A. Wolff von Gudenberg, A.L. Giraud, H. Lanfermann and V. Gall et al., (2003). The nature and treatment of stuttering as revealed by fMRI.A within- and between-group comparison, Journal of Fluency Disorders 28 (2003), pp. 381–410.

Nicholls, M. E. (1996). Temporal processing asymmetries between the cerebral hemispheres: evidence and implications. Laterality: Asymmetries of Body, Brain and Cognition, 1(2), 97-138.

Nilsonne, Å., & Sundberg, J. (1985). Differences in ability of musicians and nonmusicians to judge emotional state from the fundamental frequency of voice samples. Music Perception, 507-516.

Nooteboom, S. G. (1972). Production and perception of vowel duration: A study of durational properties of vowels in Dutch. Philips Research Laboratories.

Norris, D., & Cutler, A. (1988). The relative accessibility of phonemes and syllables. Perception & Psychophysics, 43(6), 541-550.

Norton, A., Zipse, L., Marchina, S., & Schlaug, G. (2009). Melodic intonation therapy. Annals of the New York Academy of Sciences, 1169(1), 431-436.

Nozaradan, S., Peretz, I., Missal, M., & Mouraux, A. (2011). Tagging the neuronal entrainment to beat and meter. The Journal of Neuroscience, 31(28), 10234-10240.

Näätänen, R. (1990). The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. Behavioral Brain Sciences, 13, 201–288.

Näätänen, R., Jacobsen, T., & Winkler, I. (2005). Memory- based or afferent processes in mismatch negativity (MMN): A review of the evidence. Psychophysiology, 42(1), 25-32.

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. Clinical Neurophysiology, 118(12), 2544-2590.

O'Hara, M., & Johnston, J. (1997). Syntactic bootstrapping in children with specific language impairment. International Journal of Language & Communication Disorders, 32(2), 189-205.

Ohnishi, T., Matsuda, H., Asada, T., Aruga, M., Hirakata, M., Nishikawa, M., ... & Imabayashi, E. (2001). Functional anatomy of musical perception in musicians. Cerebral Cortex, 11(8), 754-760.

Otake, T., Hatano, G., Cutler, A., & Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. Journal of Memory and Language, 32(2), 258-278.

Overy, K. (2000). Dyslexia, temporal processing and music: The potential of music as an early learning aid for dyslexic children. Psychology of Music, 28, 218–229. doi:10.1177/0305735600282010

Overy, K. (2003). Dyslexia and music. Annals of the New York Academy of Sciences, 999(1), 497-505.

Overy, K., Nicolson, R. I., Fawcett, A. J., & Clarke, E. F. (2003). Dyslexia and music: Measuring musical timing skills. Dyslexia, 9, 18–36. doi: 10.1002/dys.233

Oviatt, S. 1995. Predicting spoken disfluencies during human-computer interaction. Computer Speech and Language 9.19-35.

P. Vuust, A. Roepstorff, M. Wallentin, K. Mouridsen, L. Ostergaard, It don't meana thing. . . Keeping the rhythm during polyrhythmic tension, activates language areas (BA47), Neuroimage 31 (2006) 832–841.

Packman, Mark Onslow, Ross Menzies, A. (2000). Novel speech patterns and the treatment of stuttering. Disability & Rehabilitation, 22(1-2), 65-79.

Palmer, C., & Krumhansl, C. L. (1990). Mental representations for musical meter. Journal of Experimental Psychology: Human Perception and Performance, 16(4), 728.

Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L. E., & Hoke, M. (1998). Increased auditory cortical representation in musicians. Nature, 392(6678), 811-814.

Parbery-Clark, A., Skoe, E., Lam, C., & Kraus, N. (2009). Musician enhancement for speech-in-noise. Ear and hearing, 30(6), 653-661.

Parbery-Clark, A., Tierney, A., Strait, D. L., & Kraus, N. (2012). Musicians have fine-tuned neural distinction of speech syllables. Neuroscience, 219, 111-119.

Parncutt, R. (1994). A perceptual model of pulse salience and metrical accent in musical rhythms. Music Perception, 409-464.

Patel, A. and Balaban, E. (2001) Human pitch perception is reflected in the timing of stimulus- related cortical activity. Nat. Neurosci.
4, 839–844

Patel, A. D. (1998). Syntactic processing in language and music: different cognitive operations, similar neural resources?. Music Perception, 27-42.

Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. Frontiers in psychology, 2.

Patel, A.D. (2012). Language, music, and the brain: a resource-sharing framework. In: P. Rebuschat, M. Rohrmeier, J. Hawkins, & I. Cross (Eds.), Language and Music as Cognitive Systems (pp. 204-223). Oxford: Oxford University Press.

Patel, A. D., & Daniele, J. R. (2003). An empirical comparison of rhythm in language and music. Cognition, 87(1), B35-B45.

Patel, A.D., & Iversen, J.R. (2007). The linguistic benefits of musical abilities. Trends in Cognitive Sciences, 11, 369-372.

Patel, A. D., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. J. (1998). Processing syntactic relations in language and music: An event-related potential study. Journal of cognitive neuroscience, 10(6), 717-733.

Patel, A. D., Iversen, J. R., & Rosenberg, J. C. (2006). Comparing the rhythm and melody of speech and music: The case of British English and French. The Journal of the Acoustical Society of America, 119, 3034.

Patel, A. D., Iversen, J. R., Chen, Y., & Repp, B. H. (2005). The influence of metricality and modality on synchronization with a beat. Experimental Brain Research, 163(2), 226-238.

Patel, A. D., Wong, M., Foxton, J., Lochy, A., & Peretz, I. (2008). Speech intonation perception deficits in musical tone deafness (congenital amusia). Music Perception: An Interdisciplinary Journal, 25(4), 357-368.

Patel, A.D. 2008. Music, Language, and the Brain. Oxford 175.University Press. Oxford.

Peelle, J. E., & Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. Frontiers in psychology, 3.

Penfield W and Roberts L (1959) Speech and brain mechanisms. Princeton, NJ: Princeton University Press.

Penhune, V. B., Zatorre, R. J., & Evans, A. C. (1998). Cerebellar contributions to motor timing: a PET study of auditory and visual rhythm reproduction. Journal of cognitive neuroscience, 10(6), 752-765.

Peretz, I. (1990). Processing of local and global musical information by unilateral brain-damaged patients. Brain, 113(4), 1185-1205.

Peretz, I., & Morais, J. (1989). Music and modularity. Contemporary Music Review, 4, 277–291.

Peretz, I., Ayotte, J., Zatorre, R. J., Mehler, J., Ahad, P., Penhune, V. B., & Jutras, B. (2002). Congenital amusia: a disorder of fine-grained pitch discrimination. Neuron, 33(2), 185-191.

Peretz, I., & Coltheart, M. (2003). Modularity of music processing. Nature neuroscience, 6(7), 688-691.

Petacchi, A., Laird, A. R., Fox, P. T., & Bower, J. M. (2005). Cerebellum and auditory function: An ALE meta- analysis of functional neuroimaging studies. Human brain mapping, 25(1), 118-128.

Phillips, D.P. and Farmer, M.E. (1990) Acquired word deafness and the temporal grain of sound representation in the primary auditory cortex. Behav. Brain Res. 40, 84–90

Phillips-Silver, J., Toiviainen, P., Gosselin, N., Piché, O., Nozaradan, S., Palmer, C., & Peretz, I. (2011). Born to dance but beat deaf: A new form of congenital amusia. Neuropsychologia, 49(5), 961-969.

Pike, K.L. 1945: The intonation of American English. Ann Arbor: University of Michigan Press.

Pinker, S. (1997). How the mind works (1st ed.). New York: Norton.

Pitt, M. A., & Monahan, C. B. (1987). The perceived similarity of auditory polyrhythms. Perception & psychophysics, 41(6), 534-546.

Pitt, M. A., & Samuel, A. G. (1990a). The use of rhythm in attending to speech. Journal of Experimental Psychology: Human Perception and Performance, 16(3), 564.

Pitt, M. A., & Samuel, A. G. (1990b). Attentional allocation during speech perception: How fine is the focus?. Journal of Memory and Language, 29(5), 611-632.

Poeppel, D. (2001). Pure word deafness and the bilateral processing of the speech code. Cognitive Science, 25(5), 679-693.

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: cerebral lateralization as ‘asymmetric sampling in time’. Speech Communication, 41(1), 245-255.

Polk, M., & Kertesz, A. (1993). Music and language in degenerative disease of the brain. Brain and Cognition, 22(1), 98-117.

Port, R., Cummins, F., & Gasser, M. (1995). A dynamic approach to rhythm in language: Toward a temporal phonology. arXiv preprint cmp-lg/9508007.

Port, R. F. (2003). Meter and speech. Journal of Phonetics, 31(3), 599-611.

Port, R.F., Dalby, J. and O’Dell, M. (1987) Evidence for mora-timing in Japanese. Journal of the Acoustical Society of America 81, pp. 1574-85.

Port, R. F., Tajima, K., & Cummins, F. (1996). Self-entrainment in animal behavior and human speech. In Online proceedings of the 1996 midwest artificial intelligence and cognitive science conference.

Postma, A., & Kolk, H. (1993). The covert repair hypothesis: Prearticulatory repair processes in normal and stuttered disfluencies. Journal of Speech, Language and Hearing Research, 36(3), 472.

Potter, D. D., Fenwick, M., Abecasis, D., & Brochard, R. (2009). Perceiving rhythm where none exists: Event-related potential (ERP) correlates of subjective accenting. cortex, 45(1), 103-109.

Poulin-Charronnat, B., Bigand, E., Madurell, F., & Peereman, R. (2005). Musical structure modulates semantic priming in vocal music. Cognition, 94(3), B67-B78.

Povel, D. J., & Essens, P. (1985). Perception of temporal patterns. Music Perception, 411-440.

Preibisch et al., 2003a C. Preibisch, K. Neumann, P. Raab, H.A. Euler, A.W. von Gudenberg and H. Lanfermann et al., Evidence for compensation for stuttering by the right frontal operculum, Neuroimage 20 (2003), pp. 1356–1364

Prins, D., Hubbard, C. P ., & Krause, M. (1991). Syllabic stress and the occurrence of stuttering. Journal of Speech and HearingResearch, 34, 1011-1016.

Przybylski, L., Bedoin, N., Krifi-Papoz, S., Herbillon, V., Roch, D., Léculier, L., Kotz, S. & Tillmann, B. (2013). Rhythmic auditory stimulation influences syntactic processing in children with developmental language disorders. Neuropsychology, 27(1), 121.

Quené, H., & Koster, M. L. (1998). Metrical segmentation in Dutch: Vowel quality or stress?. Language and Speech, 41(2), 185-202.

Quené, H., & Port, R. (2005). Effects of timing regularity and metrical expectancy on spoken-word perception. Phonetica, 62(1), 1-13.

Rammsayer, T., & Altenmüller, E. (2006). Temporal information processing in musicians and nonmusicians. Music Perception: An Interdisciplinary Journal, 24(1), 37-48.

Ramus, F. (2002a). Acoustic correlates of linguistic rhythm: perspectives. Proceedings of Speech Prosody 2002, Aix-en-Provence (pp. 115–120). Aix-en-Provence: Laboratoire Parole et Langage.

Ramus, F. (2002b). Language discrimination by newborns: teasing apart phonotactic, rhythmic, and intonational cues. Annual Review of Language Acquisition, 2, 85–115.

Ramus, F., Nespor, M., & Mehler, J. (2000). Correlates of linguistic rhythm in the speech signal. Cognition, 75(1), AD3-AD30.

Rao, S. M., Harrington, D. L., Haaland, K. Y., Bobholz, J. A., Cox, R. W., & Binder, J. R. (1997). Distributed neural systems underlying the timing of movements. The Journal of Neuroscience, 17(14), 5528-5535.

Reeves, C., Schmauder, A. R., & Morris, R. K. (2000). Stress grouping improves performance on an immediate serial list recall task. Journal of Experimental Psychology: Learning, Memory, and Cognition, 26(6), 1638.

Repp, B. H. (2005). Sensorimotor synchronization: A review of the tapping literature. Psychonomic Bulletin & Review, 12(6), 969-992.

Repp, B. H., & Su, Y. H. (2013). Sensorimotor synchronization: A review of recent research (2006–2012). Psychonomic bulletin & review, 1-50.

Riley, G. (1994). Stuttering severity instrument for young children (SSI-3). Stuttering severity instrument for young children (SSI-3).

Rinne, T., Alho, K., Alku, P., Holi, M., Sinkkonen, J., Virtanen, J., Bertrand, O., Na □a □ta □nen, R., 1999. Analysis of speech sounds is left- hemisphere predominant at 100–150ms after sound onset. NeuroReport 10, 1113–1117.

Roach, P. (1982). On the distinction between ‘stress-timed’and ‘syllable-timed’languages. Linguistic controversies, 73-79.

Roncaglia-Denissen, M. P., Schmidt-Kassow, M., & Kotz, S. A. (2013). Speech rhythm facilitates syntactic ambiguity resolution: ERP evidence. PloS one, 8(2), e56000.

Rossi, Mario (1980). Le français, langue sans accent? In: Ivan Fónagy & Pierre P. Léon (eds.) L’accent en français contemporain. (Studia Phonetica 15). Montréal: Didier. 13–51.

Rothermich, K., Schmidt-Kassow, M., & Kotz, S. A. (2012). Rhythm's gonna get you: Regular meter facilitates semantic sentence processing. Neuropsychologia, 50(2), 232-244.

Rothermich, K., Schmidt-Kassow, M., Schwartze, M., & Kotz, S. A. (2010). Event-related potential responses to metric violations: rules versus meaning. Neuroreport, 21(8), 580-584.

Russo, N., Nicol, T., Trommer, B., Zecker, S., & Kraus, N. (2009). Brainstem transcription of speech is disrupted in children with autism spectrum disorders. Developmental science, 12(4), 557-567.

Russo NM, Nicol TG, Zecker SG, Hayes EA, Kraus N. Auditory training improves neural timing in the human brainstem. Behav Brain Res. 2005;156:95–103.

Sakai, K., Hikosaka, O., Miyauchi, S., Takino, R., Tamada, T., Iwata, N. K., & Nielsen, M. (1999). Neural representation of a rhythm depends on its interval ratio. The Journal of Neuroscience, 19(22), 10074-10081.

Samson, S., Ehrlé, N., & Baulac, M. (2001). Cerebral substrates for musical temporal processes. Annals of the New York Academy of Sciences, 930(1), 166-178.

Samuel AG. (1987) Phonemic restoration: insights from a new methodology. J Exp Psychol Gen. 1981;110:474–494.

Savin, H. B., & Bever, T. G. (1970). The nonperceptual reality of the phoneme. Journal of Verbal Learning and Verbal Behavior, 9, 295–302.

Schellenberg, E. G. (2004). Music lessons enhance IQ. Psychological Science, 15(8), 511-514.

Schellenberg, E. G. (2011). Examining the association between music lessons and intelligence. British Journal of Psychology, 102(3), 283-302.

Schellenberg, E. G., Krysciak, A. M., & Campbell, R. J. (2000). Perceiving emotion in melody: Interactive effects of pitch and rhythm. Music Perception, 155-171.

Schlaug, G. (2001). The brain of musicians. Annals of the New York Academy of Sciences, 930(1), 281-299.

Schlaug, G., Jancke, L., Huang, Y., & Steinmetz, H. (1995). In vivo evidence of structural brain asymmetry in musicians. Science, 267(5198), 699-701.

Schmidt-Kassow, M., & Kotz, S. A. (2009a). Event-related brain potentials suggest a late interaction of meter and syntax in the P600. Journal of Cognitive Neuroscience, 21(9), 1693-1708.

Schmidt-Kassow, M. & Kotz, S. A. (2009b) Attention and perceptual regularity in speech. NeuroReport 20:18, 1643-1647

Schmidt-Kassow, M., Rothermich, K., Schwartze, M., & Kotz, S. A. (2011). Did you get the beat? Late proficient French-German learners extract strong–weak patterns in tonal but not in linguistic sequences. Neuroimage, 54(1), 568-576.

Schmidt-Kassow, M., Schubotz, R. I., & Kotz, S. A. (2009). Attention and entrainment: P3b varies as a function of temporal predictability. Neuroreport, 20(1), 31-36.

Schneider, P., Scherg, M., Dosch, H. G., Specht, H. J., Gutschalk, A., & Rupp, A. (2002). Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians. Nature neuroscience, 5(7), 688-694.

Schneider, P., Sluming, V., Roberts, N., Scherg, M., Goebel, R., Specht, H. J., ... & Rupp, A. (2005). Structural and functional asymmetry of lateral Heschl's gyrus reflects pitch perception preference. Nature neuroscience, 8(9), 1241-1247.

Schnitzler, A., Salenius, S., Salmelin, R., Jousmäki, V., & Hari, R. (1997). Involvement of primary motor cortex in motor imagery: a neuromagnetic study. Neuroimage, 6(3), 201-208.

Schofield B. R. (2010). Projections from auditory cortex to midbrain cholinergic neurons that project to the inferior colliculus. Neuroscience 166, 231–240. doi: 10.1016/j.neuroscience.2009.12.008.

Schuppert, M., Münte, T. F., Wieringa, B. M., & Altenmüller, E. (2000). Receptive amusia: evidence for cross-hemispheric neural networks underlying music processing strategies. Brain, 123(3), 546-559.

Schwartze, M., Keller, P. E., Patel, A. D., & Kotz, S. A. (2011). The impact of basal ganglia lesions on sensorimotor synchronization, spontaneous motor tempo, and the detection of tempo changes. Behavioural Brain Research, 216, 685–691. doi:10.1016/j.bbr.2010.09.015

Schön, D., & François, C. (2011). Musical expertise and statistical learning of musical and linguistic structures. Frontiers in psychology, 2.

Schön, D., Magne, C., & Besson, M. (2004). The music of speech: Music training facilitates pitch processing in both music and language. Psychophysiology, 41(3), 341-349.

Scott, S. K. & Johnsrude, I. S. The neuroanatomical and functional organization of speech perception. Trends Neurosci. 26, 100–107 (2003).

Scott, S. K., McGettigan, C., & Eisner, F. (2009). A little more conversation, a little less action—candidate roles for the motor cortex in speech perception. Nature Reviews Neuroscience, 10(4), 295-302.

Scott, S. K. The point of P-centres. Psychol. Res. 61, 4–11 (1998).

Serafine, M. L. (1975). A measure of meter conservation in music, based on Piaget's theory (Doctoral dissertation, University of Florida).

Shahin, A., Bosnyak, D. J., Trainor, L. J., & Roberts, L. E. (2003). Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. The Journal of Neuroscience, 23(13), 5545-5552.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. Science, 270(5234), 303-304.

Shapiro, B. E., Grossman, M., & Gardner, H. (1981). Selective musical processing deficits in brain damaged populations. Neuropsychologia, 19(2), 161-169.

Shields, J. L., McHugh, A., & Martin, J. G. (1974). Reaction time to phoneme targets as a function of rhythmic cues in continuous speech. Journal of Experimental Psychology, 102(2), 250.

Shiller, D. M., Sato, M., Gracco, V. L., & Baum, S. R. (2009). Perceptual recalibration of speech sounds following speech motor learning. The Journal of the Acoustical Society of America, 125, 1103.

Shook, A., Marian, V., Bartolotti, J., & Schroeder, S. R. (2013). Musical Experience Influences Statistical Learning of a Novel Language. The American journal of psychology, 126(1), 95-104.

Shriberg, L. D., Austin, D., Lewis, B. A., McSweeny, J. L., & Wilson, D. L. (1997). The percentage of consonants correct (PCC) metric: extensions and reliability data. Journal of Speech, Language and Hearing Research, 40(4), 708.

Shtyrov, Y., Kujala, T., Palva, S., Ilmoniemi, R.J., Na a ta nen, R., 2000. Discrimination of speech and of complex nonspeech sounds of different temporal structure in the left and right cerebral hemispheres. Neuro- Image 12, 657–663.

Sices, L., Taylor, H. G., Freebairn, L., Hansen, A., & Lewis, B. (2007). Relationship between speech-sound disorders and early literacy skills in preschool-age children: impact of comorbid language impairment. Journal of developmental and behavioral pediatrics: JDBP, 28(6), 438.

Slevc, L. R. (2012). Language and music: sound, structure, and meaning. Wiley Interdisciplinary Reviews: Cognitive Science, 3(4), 483-492.

Slevc, L. R., Rosenberg, J. C., & Patel, A. D. (2009). Making psycholinguistics musical: Self-paced reading time evidence for shared processing of linguistic and musical syntax. Psychonomic bulletin & review, 16(2), 374-381.

Sluijter, A. M., & van Heuven, V. J. (1996, October). Acoustic correlates of linguistic stress and accent in Dutch and American English. In Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on (Vol. 2, pp. 630-633). IEEE.

Sluming, V., Barrick, T., Howard, M., Cezayirli, E., Mayes, A., & Roberts, N. (2002). Voxel-based morphometry reveals increased gray matter density in Broca's area in male symphony orchestra musicians. Neuroimage, 17(3), 1613-1622.

Snowling, M.J., Phonemic deficits in developmental dyslexia, Psychol. Res., 43 (1981) 219–234.

Snyder J. S., Large E. W. (2005). Gamma-band activity reflects the metric structure of rhythmic tone sequences. Cogn. Brain Res. 24, 117–126. doi: 10.1016/j.cogbrainres.2005.03.002.

Song, J. H., Skoe, E., Wong, P. C., & Kraus, N. (2008). Plasticity in the adult human auditory brainstem following short-term linguistic training. Journal of Cognitive Neuroscience, 20(10), 1892-1902.

Sowiński, J., & Dalla Bella, S. (2013). Poor synchronization to the beat may result from deficient auditory-motor mapping. Neuropsychologia.

Speedie, L. J., Wertman, E., Ta'ir, J., & Heilman, K. M. (1993). Disruption of automatic speech following a right basal ganglia lesion. Neurology, 43(9), 1768-1768.

Squires, K. C., Wickens, C., Squires, N. K., & Donchin, E. (1976). The effect of stimulus sequence on the waveform of the cortical event-related potential. Science.

Stahl, B., Henseler, I., Turner, R., Geyer, S., & Kotz, S. A. (2013). How to engage the right brain hemisphere in aphasics without even singing: evidence for two paths of speech recovery. Frontiers in human neuroscience, 7.

Stahl, B., Kotz, S. A., Henseler, I., Turner, R., & Geyer, S. (2011). Rhythm in disguise: why singing may not hold the key to recovery from aphasia. Brain, 134(10), 3083-3093.

Steinbeis, N., & Koelsch, S. (2008). Shared neural resources between music and language indicate semantic processing of musical tension-resolution patterns. Cerebral Cortex, 18(5), 1169-1178.

Steinhauer, K., Alter, K., & Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. Nature neuroscience, 2(2), 191-196.

Stewart, L., von Kriegstein, K., Warren, J. D., & Griffiths, T. D. (2006). Music and the brain: disorders of musical listening. Brain, 129(10), 2533-2553.

Strait, D., & Kraus, N. (2011). Playing music for a smarter ear: cognitive, perceptual and neurobiological evidence. Music perception, 29(2), 133.

Stromswold, K., Caplan, D., Alpert, N., Rauch, S. (1996). Localization of syntactic comprehension by positron emission tomography. Brain & Lang, 52, 452-473.

Suresh, R., Ambrose, N., Roe, C., Pluzhnikov, A., Wittke-Thompson, J. K., Ng, M. C. Y., ... & Cox, N. J. (2006). New complexities in the genetics of stuttering: significant sex-specific linkage signals. The American Journal of Human Genetics, 78(4), 554-563.

Tajima, K. (1998). Speech rhythm in English and Japanese: Experiments in speech cycling (Doctoral dissertation, Indiana University).

Tallal, P., Miller, S. L., Bedi, G., Byma, G., Wang, X., Nagarajan, S. S., ... & Merzenich, M. M. (1996). Language comprehension in language-learning impaired children improved with acoustically modified speech. Science, 271(5245), 81-84.

Teki, S., Grube, M., Kumar, S., & Griffiths, T. D. (2011). Distinct neural substrates of duration-based and beat-based auditory timing. The Journal of Neuroscience, 31(10), 3805-3812.

Thaut, M. (2005). Rhythm music and the brain: Scientific foundations and clinical applications (Vol. 7). Routledge.

Thaut, M. H., & Abiru, M. (2010). Rhythmic auditory stimulation in rehabilitation of movement disorders: a review of current research. Music Perception: An Interdisciplinary Journal, 27(4), 263-269.

Thaut, M. H., Kenyon, G. P., Schauer, M. L., & McIntosh, G. C. (1999). The connection between rhythmicity and brain function. Engineering in Medicine and Biology Magazine, IEEE, 18(2), 101-108.

Thaut, M. H., McIntosh, G. C., Rice, R. R., Miller, R. A., Rathbun, J., & Brault, J. M. (1996). Rhythmic auditory stimulation in gait training for Parkinson's disease patients. Movement Disorders, 11(2), 193-200.

Thompson, W. F., Schellenberg, E. G., & Husain, G. (2003). Perceiving prosody in speech: Effects of music lessons. Annals of the New York Academy of Sciences, 999, 530–532.

Thompson, W.F., Schellenberg, E.G., & Husain, G. (2004). Decoding speech prosody: Do music lessons help? Emotion, 4, 46-64.

Thomson, J. M., & Goswami, U. (2008). Rhythmic processing in children with developmental dyslexia: Auditory and motor rhythms link to reading and spelling. Journal of Physiology-Paris, 102(1), 120-129.

Tierney, A., & Kraus, N. (2013a). Neural responses to sounds presented on and off the beat of ecologically valid music. Frontiers in systems neuroscience, 7.

Tierney, A., & Kraus, N. (2013b). The Ability to Move to a Beat Is Linked to the Consistency of Neural Responses to Sound. The Journal of Neuroscience, 33(38), 14981-14988.

Tillmann, B. (2012). Music and language perception: expectations, structural integration, and cognitive sequencing. Topics in cognitive science, 4(4), 568-584.

Tillmann, B., & Lebrun-Guillaud, G. (2006). Influence of tonal and temporal expectations on chord processing and on completion judgments of chord sequences. Psychological Research, 70(5), 345-358.

Tillmann, B., Janata, P., & Bharucha, J. J. (2003). Activation of the inferior frontal cortex in musical priming. Cognitive Brain Research, 16, 145-161.

Toyomura, A., Fujii, T., & Kuriki, S. (2011). Effect of external auditory pacing on the neural activity of stuttering speakers. NeuroImage, 57(4), 1507-1516.

Trainor, L. J., McDonald, K. L., & Alain, C. (2002). Automatic and controlled processing of melodic contour and interval information measured by electrical brain activity. Journal of Cognitive Neuroscience, 14(3), 430-442.

Trajkovski, N., Andrews, C., Onslow, M., Packman, A., O'Brian, S., & Menzies, R. (2009). Using syllable-timed speech to treat preschool children who stutter: A multiple baseline experiment. Journal of fluency disorders, 34(1), 1-10.

Tueting, P., Sutton, S., & Zubin, J. (1970). Quantitative evoked potential correlates of the probability of events. Psychophysiology, 7(3), 385-394.

Vaissière, J. (1974). On French prosody. Quaterly Progress Report, Research Laboratory of Electronics, Massachusetts Institute of Technology, (114), 212-223.

Vaissière, J. (1983). Language-Independent Prosodic Features. In: Anne Cut- ler & D. Robert Ladd (eds.) Prosody: Models and Measurements. Berlin: Springer Verlag. 53–66.

Vaissière, J. (1991). Rhythm, accentuation and final lengthening in French. Music, language, speech and brain, 108-120.

van Ede, F., de Lange, F., Jensen, O., & Maris, E. (2011). Orienting attention to an upcoming tactile event involves a spatially and temporally specific modulation of sensorimotor alpha- and beta-band oscillations. The Journal of Neuroscience, 31(6), 2016-2024..

Van Lancker Sidtis, D., Cameron, K., & Sidtis, J. J. (2012). Dramatic effects of speech task on motor and linguistic planning in severely dysfluent parkinsonian speech. Clinical linguistics & phonetics, 26(8), 695-711.

van Noorden, L., & Moelants, D. (1999). Resonance in the perception of musical pulse. Journal of New Music Research, 28(1), 43-66.

Van Zuijen, T. L., Sussman, E., Winkler, I., Näätänen, R., & Tervaniemi, M. (2004). Grouping of sequential sounds—an event-related potential study comparing musicians and nonmusicians. Journal of cognitive neuroscience, 16(2), 331-338.

Vargha-Khadem, F., Watkins, K. E., Price, C. J., Ashburner, J., Alcock, K. J., Connelly, A., ... & Passingham, R. E. (1998). Neural basis of an inherited speech and language disorder. Proceedings of the National Academy of Sciences, 95(21), 12695-12700.

Velasco, M. J., & Large, E. W. (2011). Pulse detection in syncopated rhythms using neural oscillators. pulse, 1(2), 3-4.

VIGNOLO, L. A. (2003). Music agnosia and auditory agnosia. Annals of the New York Academy of Sciences, 999(1), 50-57.

Volkmann, J., Hefter, H., Lange, H. W., & Freund, H. J. (1992). Impairment of temporal organization of speech in basal ganglia diseases. Brain and Language, 43(3), 386-399.

Vuust, P., Pallesen, K. J., Bailey, C., van Zuijen, T. L., Gjedde, A., Roepstorff, A., & Østergaard, L. (2005). To musicians, the message is in the meter: pre-attentive neuronal responses to incongruent rhythm are left-lateralized in musicians. Neuroimage, 24(2), 560-564.

Wallentin, M., Nielsen, A. H., Friis-Olivarius, M., Vuust, C., & Vuust, P. (2010). The Musical Ear Test, a new reliable test for measuring musical competence. Learning and Individual Differences, 20(3), 188-196.

Wallin, L., Merker, B., & Brown, S. (Eds.). (2001). The origins of music. The MIT Press.

Wang, X.F. et al. (1991) The dentate nucleus is a short-latency relay of a primary auditory transmission pathway. Neuroreport 2, 361–364

Ward, L. M. (2003). Synchronous neural oscillations and cognitive processes. Trends in cognitive sciences, 7(12), 553-559.

Watkins, K. E., Strafella, A. P., & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. Neuropsychologia, 41(8), 989-994.

Weber, C., Hahne, A., Friedrich, M., & Friederici, A. D. (2004). Discrimination of word stress in early infant perception: Electrophysiological evidence. Cognitive Brain Research, 18(2), 149-161.

Webster, R. L. (1980). Evolution of a target-based behavioral therapy for stuttering. Journal of Fluency Disorders, 5(3), 303-320.

Weinert, S. (1992). Deficits in acquiring language structure: The importance of using prosodic cues. Applied Cognitive Psychology, 6, 545–571. doi:10.1002/acp.2350060607

Wenk, B. J. (1987). Just in time: on speech rhythms in music. Linguistics, 25, 969–981.

Wildgruber, D., Ackermann, H., & Grodd, W. (2001). Differential contributions of motor cortex, basal ganglia, and cerebellum to speech motor control: effects of syllable repetition rate evaluated by fMRI. Neuroimage, 13(1), 101-109.

Wilson, S. J., Pressing, J. L., & Wales, R. J. (2002). Modelling rhythmic function in a musician post-stroke. Neuropsychologia, 40(8), 1494-1505.

Wingate, M. E. (1976). Stuttering: Theory and treatment. New York: Irvington.

Wingate, M. E. (1988). The structure of stuttering: A psycholinguistic perspective. New York:Springer-Verlag.

Winkler, I., Háden, G. P., Ladinig, O., Sziller, I., & Honing, H. (2009). Newborn infants detect the beat in music. Proceedings of the National Academy of Sciences, 106(7), 2468-2471.

Wong, P. C., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. Nature neuroscience, 10(4), 420-422.

Wong, P. C., Warrier, C. M., Penhune, V. B., Roy, A. K., Sadehh, A., Parrish, T. B., & Zatorre, R. J. (2008). Volume of left Heschl's gyrus and linguistic pitch learning. Cerebral Cortex, 18(4), 828-836.

Woodrow, H. H. (1909). A quantitative study of rhythm: The effect of variations in intensity, rate and duration (No. 14). Science Press.

Yairi, E. (2007). Subtyping stuttering I: a review. Journal of fluency disorders, 32(3), 165-196.

Yairi, E., & Ambrose, N. G. (1999). Early childhood stuttering I: persistency and recovery rates. J Speech Lang Hear Res, 42(5), 1097–1112.

Yairi, E., & Ambrose, N. G. (2005). Early childhood stuttering for clinicians by clinicians. Pro Ed.

Yairi, E., Ambrose, N. G., Paden, E. P., & Throneburg, R. N. (1996). Predictive factors of persistence and recovery: Pathways of childhood stuttering. Journal of Communication Disorders, 29(1), 51-77

Yee, W., Holleran, S., & Jones, M. R. (1994). Sensitivity to event timing in regular and irregular sequences: Influences of musical skill. Perception & Psychophysics, 56(4), 461-471.

Zanto, T. P., Snyder, J. S., & Large, E. W. (2006). Neural correlates of rhythmic expectancy. Advances in Cognitive Psychology, 2(2-3), 221-231.

Zatorre, R., Mehler, J., Ahad, P., Pen- hune, V. B., & Jutras, B. (2002). Congenital amusia: A disorder of fine-grained pitch discrimination. Neuron, 33, 185–191.

Zatorre, R.J. (1988) Pitch perception of complex tones and human temporal-lobe function. J. Acoust. Soc. Am. 84, 566–572

Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: music and speech. Trends in cognitive sciences, 6(1), 37-46.

Zatorre, R.J., Belin, P., (2001). Spectral and temporal processing in human auditory cortex. Cereb. Cortex 11, 946–953.

Zatorre, R.J., Belin, P., Penhune, V.B., 2002. Structure and function of auditory cortex: music and speech. Trends Cognit. Sci. 6, 37–46.

Zatorre, R. J., Chen, J. L., & Penhune, V. B. (2007). When the brain plays music: auditory–motor interactions in music perception and production. Nature Reviews Neuroscience, 8(7), 547-558.

Zatorre, R. J., Perry, D. W., Beckett, C. A., Westbury, C. F., & Evans, A. C. (1998). Functional anatomy of musical processing in listeners with absolute pitch and relative pitch. Proceedings of the National Academy of Sciences, 95(6), 3172-3177.

Zelaznik HN, Spencer RMC, Ivry RB (2004): Dissociation of explicit and implicit timing in repetitive tapping and drawing movements, J Exp Psychol Hum Percept Perform 28:575-588.

Zellner, B. (1994). Pauses and the temporal structure of speech. Zellner, B.(1994). Pauses and the temporal structure of speech, in E. Keller (Ed.) Fundamentals of speech synthesis and speech recognition.(pp. 41-62). Chichester: John Wiley., 41-62.

Zentner, M. & T. Eerola. 2010. Rhythmic engagement with music in infancy. Proc. Natl. Acad. Sci. USA 107: 5768–5773.

Zion Golumbic, E. M., Poeppel, D., & Schroeder, C. E. (2012). Temporal context in speech processing and attentional stream selection: a behavioral and neural perspective. Brain and language, 122(3), 151-161.

TITRE en français: **L’effet du rythme musical sur la parole**

RESUME en français :

La musique et la parole reposent sur une organisation temporelle. En effet, l’anticipation, l’organisation et le groupement temporel y sont nécessaires. Il serait donc possible que des processus domaine-généraux sous-tendent ces deux processus temporels soient mis en jeu.

Pour tester cette hypothèse, trois expériences utilisant des mesures comportementales et électrophysiologiques (EEG) ont été menées afin de déterminer si la perception et la production de la parole peuvent bénéficier d’un amorçage rythmique (i.e. la présentation au préalable d’un rythme musical qui peut renseigner sur les structures temporelles de la parole). En utilisant ces mesures, nous avons montré que le traitement phonologique de pseudo-mots parlés est renforcé lorsque la parole est conforme aux prédictions temporelles des auditeurs (Cason & Schön, 2012). Le traitement phonologique des phrases peut également être amélioré et cet effet d’amorçage augmenté grâce à un entrainement avec les rythmes musicaux (Cason, Astésano & Schön, soumis). Dans une troisième étude, nous avons montré que l’amorçage rythmique peut augmenter la production phonologique chez les enfants sourds (Cason, Hidalgo & Schön, soumis).

Ces trois études montrent que la régularité du rythme musical (plus important que dans la parole) semblerait permettre, de manière générale, la formation de prédictions temporelles précises et une trace mnésique également améliorer la production de la parole chez les enfants souffrants de troubles auditif.

TITRE en anglais: **The effect of musical rhythm on spoken language**

RESUME en anglais:

Music and speech are both reliant on how events occur in time. Both require anticipation about when and what events will occur as well as a temporal and hierarchical organisation of salient and less salient events. These may rely on common, domain-general processes.

With this in mind, three experiments using behavioural and electrophysiological (EEG) measures were conducted which aimed to investigate whether speech perception and production can benefit from rhythmic priming (inducing temporal expectations through music, and which can inform a listener about temporal structures in speech). We have found that phonological processing of spoken pseudowords is enhanced when speech conforms to listener expectations, as measured by behavioural (reaction time) and EEG data (Cason & Schön, 2012). Phonological processing of sentences can also be enhanced via rhythmic priming (behavioural measures) and this priming effect is augmented through training with the musical rhythms (Cason, Astésano & Schön, submitted).

Overall, it seems that the regularity of musical rhythm (over speech rhythm) allows a listener to form precise temporal expectations and a metrical memory trace which can impact on phonological processing of words and sentences, and that rhythmic priming can also enhance articulation performance in hearing-impaired children, perhaps via an enhanced phonological perception.

MOTS-CLES/KEYWORDS
Parole, Rythme, Musique, Prédictions temporelles / Speech, Rhythm, Music, Temporal prediction