Normandie Université

UNIVERSITÉ DE ROUEN NORMANDIE

THÈSE

Pour obtenir le diplôme de doctorat

Spécialité **AUTOMATIQUE, SIGNAL, PRODUCTIQUE, ROBOTIQUE**

Préparée au sein de l'**Université de Rouen Normandie**

Représentation multimodale d'images d'oeuvres d'art

Présentée et soutenue par

DJOUSSOUF LILIA

Thèse soutenue le 01/12/2025

devant le jury composé de :

M. EL MOATAZ BILLAH ABDERRAHIM	Professeur des Universités - UCN - Université de Caen Normandie	Directeur de thèse
M. ARCHAMBAULT DOMINIQUE	Professeur des Universités - UNIVERSITE PARIS 8 UNIVERSITE VINCENNES ST DENIS	Président du jury
MME PANEELS SABRINA	Chercheur - Université Paris-Saclay	Membre du jury
M. QUEAU YVAIN	Chargé de Recherche - UCN - Université de Caen Normandie	Membre du jury
MME ROMEO KATERINE	Maître de Conférences - URN - Université de Rouen Normandie	Membre du jury
MME TREUILLET SYLVIE	Professeur des Universités - Université d'Orléans	Rapporteur du jury

Thèse dirigée par **EL MOATAZ BILLAH ABDERRAHIM** (LABORATOIRE D'INFORMATIQUE DE TRAITEMENT DE L'INFORMATION ET DES SYSTEMES)

MIIS

litis

Abstract

This thesis is part of the ANR Inclusive Museum Guide (IMG) project and focuses on access to artworks for blind and partially sighted people. While contemporary museums largely rely on vision as the primary channel of cultural mediation, blind and partially blind visitors still face numerous barriers to autonomous access, understanding, and appreciation of artworks. This audience, highly heterogeneous in their visual capacities (ranging from total blindness to residual forms of vision) mobilizes specific perceptual, cognitive, and sensory strategies to explore their environment. Museum inclusivity therefore requires rethinking mediation devices through multimodal approaches that combine tactile perception, auditory feedback, and descriptive narration, in order to offer an equitable and shared cultural experience.

The study takes as a case in point the Bayeux Tapestry, a nearly 70-meter-long medieval embroidery renowned for its narrative richness and historical significance. Its fragility and conservation constraints prevent any direct handling, making access even more challenging for blind and partially blind visitors. This context provides an ideal field for investigating virtual and sensory modalities of access to heritage artworks.

Three research questions structure this work :

— Identify perceptible information and relevant sensory channels for understanding a visual artwork ;

— Design an accessible virtual presentation adapted to different profiles of visual impairment ;

— Define methods for extracting and hierarchizing salient elements of an artwork for haptic and auditory transposition.

To this extent, the thesis contributions are organized around three main components :

— a review of existing devices, sensory modalities (tactile, auditory, haptic, verbal), and automated approaches for generating accessible representations for blind and partially blind users ;

— the development and evaluation of a 2D haptic prototype based on the F2T force-feedback tablet, exploring a hierarchical and progressive strategy for visual content exploration ;

— the design of a method for extracting salient elements from the Bayeux Tapestry, based on visual saliency data collected from sighted participants, to inform tactile and auditory adaptations.

The conceptual foundations of a processing chain for adapting an image from a visual artwork to a multisensory representation have been established. Some concrete approaches for the future design of an integrated process for generating accessible content are proposed. By encouraging autonomy, understanding, and dialogue among blind, partially blind, and sighted visitors, it supports

the creation of a truly inclusive museum experience, grounded in equitable access to cultural heritage.

Keywords : Accessibility of museum artworks ; Multimodality ; Human–Computer Interaction ; Haptic device ; Eye-tracking.

Résumé

Cette thèse s'inscrit dans le cadre du projet ANR Guide Muséal Inclusif (IMG) et s'intéresse à l'accès aux œuvres d'art pour les personnes aveugles et partiellement aveugles. Alors que les musées contemporains reposent largement sur la vision comme vecteur principal de médiation culturelle, les visiteurs aveugles ou partiellement aveugles rencontrent encore de nombreux obstacles à l'accès autonome, à la compréhension et à l'appréciation des œuvres. Ce public, très hétérogène dans leurs capacités visuelles (allant de la cécité totale à des formes résiduelles de vision), mobilisent des stratégies perceptives, cognitives et sensorielles spécifiques pour explorer leur environnement. L'inclusivité muséale suppose donc de repenser les dispositifs de médiation en intégrant des approches multimodales, combinant perception tactile, retour auditif et narration descriptive, afin d'offrir une expérience culturelle équitable et partagée.

L'étude prend pour cas emblématique la Tapisserie de Bayeux, une broderie médiévale de près de 70 mètres de long, célèbre pour sa richesse narrative et son importance historique. Sa fragilité et les contraintes de conservation interdisent toute manipulation directe, rendant son accès encore plus complexe pour les visiteurs aveugles ou partiellement aveugles. Ce contexte en fait un terrain d'étude privilégié pour questionner les modalités d'accès virtuel et sensoriel aux œuvres patrimoniales.

Trois questions de recherche structurent ce travail :

— Identifier les informations perceptibles et les canaux sensoriels pertinents pour la compréhension d'une œuvre visuelle ;
— Concevoir une présentation virtuelle accessible et adaptée à différents profils de déficience visuelle ;
— Définir des méthodes d'extraction et de hiérarchisation des éléments saillants d'une œuvre en vue d'une transposition haptique et auditive.

Pour répondre à ces interrogations, les contributions de la thèse s'articulent autour de trois volets principaux :

— un état de l'art des dispositifs existants, des modalités sensorielles exploitées (tactile, auditif, haptique, verbal) et des approches d'automatisation de la création de représentations accessibles pour les personnes aveugles et partiellement aveugles ;
— le développement et l'évaluation d'un prototype haptique 2D basé sur la tablette à retour de force F2T, explorant notamment une stratégie d'exploration hiérarchique et progressive des contenus visuels ;
— la mise au point d'une méthode d'extraction d'éléments saillants de la Tapisserie de Bayeux, à partir de données de saillance visuelle collectées auprès de personnes non aveugles, en

vue d'une adaptation tactile et auditive pertinente.

Les fondements conceptuels d'une chaîne de traitement pour l'adaptation d'une image d'œuvre visuelle à une représentation multisensorielle a été réalisée. Des pistes concrètes pour la conception future d'un processus intégré de génération de contenus accessibles sont proposées. En favorisant l'autonomie, la compréhension et le dialogue entre visiteurs aveugles, partiellement aveugles et non aveugles, cette recherche participe à la construction d'une expérience muséale véritablement inclusive, fondée sur l'équité d'accès au patrimoine culturel.

Mots-clés : Accessibilité des œuvres muséales ;Multimodalité ; Interaction Homme-Machine ; Dispositif Haptique ; Oculométrie.

Remerciements

Cette thèse n'aura pas été de tout repos, et j'ai souvent été face à des sentiments partagés tout au long de celle-ci. Je souhaite donc remercier les personnes qui ont encadré ce travail avec beaucoup d'efforts : Mme Katerine Romeo, Mme Christèle Lecomte et M. Abderrahim El Moataz.Un merci tout particulier à Katerine Romeo, sans qui cette thèse n'aurait sans doute jamais pu aboutir, pour le temps que vous avez consacré pour m'aider tout au long de celle-ci.

J'ai eu la chance de travailler sur un sujet passionnant, qui m'a permis de rencontrer des personnes engagées, bienveillantes et inspirantes. Leur regard m'a profondément sensibilisée et enrichie à travers nos échanges. Je souhaite ainsi remercier Marion Chottin, Hannah Thompson, Marjorie Redon et Alison Eardley.

J'exprime également ma gratitude à toutes les personnes du laboratoire LITIS qui ont su me conseiller, tant sur le plan technique que logistique. Merci à tous les collègues doctorants, ingénieurs et stagiaires avec qui j'ai pu sympathiser au fil de ces années, et qui m'ont permis d'avancer dans ce projet. Merci à Wenqi Luo, Nicolas Charafe et Ilies Mohammed-Riad Zizi. Merci aussi à toutes les nouvelles rencontres que j'ai pu faire, malgré ma timidité, qui ont parfois donné lieu à des moments plutôt cocasses.

Je remercie mes amis, qui m'ont aidée à me changer les idées, à me sentir soutenue et à donner le meilleur de moi-même. Merci Sabrina et Yohan pour toutes ces moments tout simplement formidables. Vous ne pouvez imaginer à quel point ces moments étaient magiques pour moi. Merci Fatima pour ta joie de vivre, ton honnêteté, et toutes les histoires croustillantes que tu sais si bien raconter et qui me changent les idées. Merci Jieyeon, Shelly, Mylène et Yasmine d'avoir toujours tendu l'oreille à mes confidences et inquiétudes, et de m'avoir offert vos conseils ainsi que l'envie de découvrir de nouveaux horizons.

Enfin, un immense merci à ma famille, dont le soutien a grandement facilité mon quotidien. Merci du fond du cœur, Maman et Papa. Merci à mes frères et sœurs Nakad, Najia, Nazim, Yamina et Nazla pour votre aide et votre présence à chaque instant. Papa, merci pour ton travail acharné, ton soutien, et toutes les inquiétudes que tu exprimes qui sont toujours sincères. Maman, tu as tant fait pour m'aider à devenir indépendante. Tu as travaillé si dur pour que je puisse suivre ma propre voie. J'ai tant de choses sur lesquelles te remercier : merci pour tous tes petits gestes, merci pour les délicieux plats que tu cuisines toujours avec amour. Et enfin, Nazim, mon petit frère, mon plus grand merci te revient, car même si tu ne peux pas t'exprimer, tu restes à chaque instant dans mes pensées. Tu ne sauras probablement jamais à quel point tu me pousses à donner le meilleur de moi-même, mais merci de tout mon cœur.

Table des matières

1 Introduction **13**

1 Contexte . . . 13

2 Questions de recherche . . . 14

3 Contributions . . . 15

4 Présentation globale de la thèse . . . 15

2 Contexte scientifique **16**

1 Introduction du chapitre . . . 16

2 Approches technologiques et sensorielles pour l'adaptation des œuvres d'art . . . 17

2.1 Dispositifs . . . 18

2.2 Méthodes d'exploration . . . 26

2.3 Hiérarchisation des informations extraites . . . 28

3 Approches pour la conception et l'adaptation de représentations d'œuvres d'art . . . 31

3.1 Co-création de représentations et segmentation interactive . . . 31

3.2 Chaînes de traitement automatiques . . . 32

3.3 Les mécanismes d'attention visuelle pour l'adaptation . . . 35

4 Synthèse . . . 36

5 Conclusion du chapitre . . . 41

3 Evaluation de l'expérience muséale **42**

1 Introduction du chapitre . . . 42

2 Étude exploratoire 1 : Définition des besoins . . . 44

2.1 Introduction . . . 44

2.2 Profil des participants . . . 45

2.3 Méthode et matériel . . . 46

2.4 Résultats . . . 49

2.5 Conclusion 53
3 Présentation du dispositif haptique : F2T 56
3.1 Description du prototype 56
3.2 Description de l'interface 57
3.3 Définition des effets 57
3.4 Modes d'interaction 59
4 Etude exploratoire 2 : Interaction avec un dispositif haptique 63
4.1 Introduction 63
4.2 Profil des participants 63
4.3 Méthode et matériel 64
4.4 Résultats & discussion 67
4.5 Conclusion 72
5 Conclusion du chapitre 73

4 Étude de la perception d'images de la Tapisserie de Bayeux 74

1 Introduction du chapitre 75
2 Base de données existantes 76
3 Base de données de la Tapisserie de Bayeux 78
3.1 Profil des participants 78
3.2 Conditions expérimentales 78
3.3 Résultats de la collecte de données 80
3.4 Discussion 95
4 Sélection de modèles 97
4.1 Introduction 97
4.2 Comparaison des méthodes de génération de cartes de saillance 97
4.3 Comparaison des méthodes de génération de parcours visuels 102
5 Conclusion du chapitre 106

5 Conclusion 109

1 Conclusion de la thèse 109
2 Perspectives 110

Table des figures

2.1 Peinture Nave Nave Mahana de Paul Gauguin . . . 21

2.2 Représentation thermogonflée d'un cheval. . . . 22

2.3 Système TACTOS. Image tirée de Rovira *et al.* [96]. . . . 27

2.4 Méthodes d'exploration identifiées pour le système TACTOS . . . 27

2.5 Méthodes d'exploration par attributs (a) et hiérarchique (b) par lecteur d'écran par Ahmetovic *et al.* [42] . . . 28

2.6 Exemple de rendu des textures tirées de Gioia *et al.* [111] . . . 33

2.7 Chaîne de traitement pour générer des images haptiques proposée par Vasudevan *et al.*. Image tirée de [130]. . . . 33

2.8 Présentation de la méthode Pic2Tac avec des images tirées de Pakenaite *et al.*[150] 35

3.1 Conditions expérimentales A, B et C pour les scènes 25, 39 et 55 issues de la Tapisserie de Bayeux . . . 47

3.2 Prototypes de la F2T . . . 56

3.3 Interface de la F2T . . . 57

3.4 Système du joystick permettant de récupérer le mouvement engendré par le doigt [169] . . . 58

3.5 Présentation de l'interface pour créer un effet de flux . . . 60

3.6 Présentation hiérarchique de différents modes d'exploration d'une image virtuelle avec la F2T . . . 60

3.7 Schéma illustratif du mode d'exploration guidée globale . . . 61

3.8 Schéma illustratif du mode d'exploration guidée détaillée . . . 61

3.9 Schéma illustratif du mode d'exploration sans guidage . . . 62

3.10 Scènes de la Tapisserie de l'Apocalypse testées . . . 65

3.11 Boîtes englobantes des éléments les plus saillants des scènes S1 et S2 . . . 65

3.12 Positionnement des attracteurs pour les scènes S1 et S2 . . . 66

3.13 Visualisation des zones audio-haptiques pour chaque scène étudiée . . . 67

3.14 Visualisation des zones de flux pour chaque scène étudiée 67
3.15 Résultats échelle de Likert pour les questions du tableau 3.9 en lien avec l'Exploration Guidée Globale (EGG), l'Exploration Guidée Détaillée (EGD), l'Exploration Sans Guidage (ESG) 68
3.16 Trajectoires des participants pour l'exploration de la scène S1 (haut) et la scène S2 (bas) 70
3.17 Estimation de densité non paramétrique par noyau de la scène S1 et la scène S2 . 71
3.18 Temps normalisé par zone audio-haptique explorée. 71

4.1 Tableau d'images illustrant les jeux de données Salicon [178], MIT1003 [179] et Le Meur [181]. 76
4.2 Tableau d'images illustrants le jeu de données OSIE [180] pour 4 participants . . 77
4.3 Configuration expérimentale avec des exemples d'images affichées 78
4.4 Présentation hiérarchique de la scène 30 de la Tapisserie de Bayeux avec descriptions verbales. 79
4.5 Tableau d'images illustrant la dispersion des points de fixations dans les zones d'intérêts 80
4.6 Tableau de comparaison des cartes de saillance pour des images de la Tapisserie de Bayeux 81
4.7 Tableau de comparaison sur des cartes de saillance pour une image précise de la Tapisserie de Bayeux 81
4.8 Observation en détails des zones saillantes pour la scène 1 de la Tapisserie de Bayeux 82
4.9 Somme cumulée des points de fixations pour chaque catégorie pour les conditions sonores : Avec et Sans audio 83
4.10 Tableau d'images de parcours visuels lors de l'observation d'une scène de la Tapisserie de Bayeux, avec et sans audio. 84
4.11 Partitionnement des images selon la méthode des K-moyennes pour trois clusters définis à partir du ratio sur la largeur (W) et la hauteur (H). 85
4.12 Exemples d'images appartenant aux cluster 0, 1 et 2. 85
4.13 Histogramme circulaire sur les directions des parcours visuels selon la condition de présence de retour audio, exprimés en fréquences relatives. 86
4.14 Histogrammes circulaires des directions des parcours visuels selon la condition de format de l'image, exprimés en fréquences relatives. 87
4.15 Boîtes à moustache des résultats pour la mesure DTW 90
4.16 Boîtes à moustache des résultats pour la métrique Multimatch 92

4.17 Résultats pour l'échelle de Likert dans le cadre des retours de l'étude sur l'observation des images de la Tapisserie de Bayeux. 95
4.18 Architecture de SAM-LSTM-ResNet[196] 98
4.19 Architecture TranSalNet-ResNet [195] 99
4.20 Tableau de comparaison des cartes de saillance pour différents modèles d'apprentissage profond 100
4.21 Architecture du modèle de Chen [198] 102
4.22 Architecture de SPGen[199] 103

Liste des tableaux

2.1 Synthèse des modalités sensorielles utilisées en fonction de l'information de l'image d'œuvre d'art 2D 29

2.2 Synthèse sur les représentations d'œuvres d'art 2D telles que les peintures et les broderies pour l'accessibilité des personnes aveugles ou partiellement aveugles selon les informations de haut et bas niveau à extraire 39

3.1 Classification de la perception visuelle selon l'OMS 45

3.2 Tableau présentant le profil visuel des participants 46

3.3 Énoncés des questions pour l'étude exploratoire 1 48

3.4 Résultats de l'échelle de Likert 49

3.5 Résultats de l'échelle de Likert 49

3.6 Résultats de l'échelle de Likert 50

3.7 Tableau présentant le profil visuel des participants ; PA : personne aveugle, PNA : personne non aveugle, PPA : personne partiellement aveugle ; F : genre féminin, M : genre masculin 64

3.8 Tableaux comportant les effets présents dans les scènes S1 et S2. A : Audio, H : Haptique 66

3.9 Enoncés liés au questionnaire de l'étude exploratoire 2. 68

4.1 Tableau des valeurs minimales et maximales des ratios calculés en largeur (W) et en hauteur (H) pour les images affichées dans une résolution de 1920×1080 pixels 86

4.2 Résultats issu du modèle bayésien pour étudier l'influence des conditions sur la similarité des observations par différents participants par la mesure DTW. La valeur de référence des modèles présentés sont des images appartenant au cluster 0 avec l'accompagnement de retours audios. 91

4.3 Résultats issus des modèles bayésiens pour étudier l'influence des conditions sur la similarité des observations par différents participants par la métrique Multimatch. La valeur de référence des modèles présentés sont des images appartenant au cluster 0 accompagnées d'audio. 94

4.4 Métriques de saillance pour les modèles testés 101
4.5 Tableau de comparaison de parcours visuels générés à partir de différents modèles . 104
4.6 Métriques pour les parcours visuels pour les modèles testés 105

1 Introduction

Sommaire

1 Contexte 13
2 Questions de recherche 14
3 Contributions 15
4 Présentation globale de la thèse 15

Cette thèse s'inscrit dans le cadre du projet ANR [1] IMG (Guide Muséal Inclusif), les principaux objectifs sont de rendre accessibles l'espace muséal et les œuvres d'arts aux personnes aveugles et partiellement aveugles. En parlant d'inclusivité, nous cherchons à développer des technologies qui seront utilisées par des personnes aveugles ou partiellement aveugles, mais aussi les personnes non aveugles.

1 Contexte

Les personnes aveugles et partiellement aveugles veulent pouvoir se déplacer librement, en sécurité et en autonomie, dans des lieux connus et inconnus. Elles sont confrontées à diverses problématiques liées à l'accessibilité des espaces sociaux. Dans le cas particulier des musées, les visites s'adressent principalement aux personnes non aveugles, et n'offrent pas ou peu d'œuvres à toucher. Depuis la loi de 2005 "Pour l'égalité des droits et des chances, la participation et la citoyenneté des personnes handicapées", les musées doivent offrir l'accessibilité à tous.

Pour le conservateur du musée, l'adaptation de l'œuvre consiste à établir quels éléments doivent être mis en avant et comment choisir l'ordre de présentation afin qu'ils soient compréhensibles et perceptibles par le toucher.

Les musées vont s'orienter vers la manipulation d'objets 3D tels que des reproductions de sculptures. Les moyens sont plus limités pour les œuvres d'art principalement 2D, telles que les peintures, les tapisseries et les broderies.

Les experts en transcription de documents manuscrits, images, photos, peintures ou tapisseries vont se tourner vers des représentations tactiles [1]. Cependant ces types de représentation peuvent engendrer des coûts élevés dans la production, et nécessitent un accompagnement par une personne qui voit [2, 3].

Il est important de se diriger vers la conception de dispositifs qui s'ancrerait dans une solution d'inclusivité lors de la conception de dispositifs pour des personnes aveugles ou partiellement aveugles.

1. Agence Nationale de la Recherche

Il apparaît souhaitable de développer des dispositifs qui, au-delà d'éviter toute forme de stigmatisation des publics concernés, soient conçus dans la perspective d'un usage partagé par des personnes aveugles et non aveugles. Une telle orientation permet non seulement de favoriser la circulation et l'enrichissement des expériences entre différents utilisateurs, mais également de limiter le risque que ces technologies demeurent sous-utilisées ou inemployées [4, 5].

Nous nous intéressons en premier lieu à l'œuvre magistrale du musée de Bayeux en Normandie. La Tapisserie de Bayeux est une broderie d'art médiéval, de 70m de long datant du XIème siècle, semblable à une bande dessinée par sa narration. Elle est protégée pour sa conservation derrière une vitrine qui la rend inaccessible aux personnes aveugles. Elle raconte l'histoire de la conquête de l'Angleterre par Guillaume le Conquérant et contient de nombreux éléments répartis sur 58 scènes.

Seules trois scènes sont représentées en relief, à l'étage, et elles n'ont pas de description audio. Dans ce contexte, l'objectif est de rendre les œuvres d'art accessibles par des sens autres que la vue, à travers des retours audios et tactiles.

Le sens du toucher correspond à une perception des objets à travers un contact cutanée mettant en action différents mécanorécepteurs. Le sens haptique, quant à lui, englobe la perception passive, mais aussi active des objets. Il est lié notamment à la notion de proprioception qui correspond à la perception de mouvements par le corps.

Bien que des règles aient été établies pour créer des graphiques tactiles par des fondations pour les aveugles (American Foundation for the Blind [2], l'INSEI (Institut national supérieur formation et recherche - handicap et enseignements adaptés) [3], l'INJA (Institut National des Jeunes Aveugles) [4]), l'adaptation de ces règles à un dispositif haptique basé sur un retour de force n'a pas encore été réalisée.

2 Questions de recherche

À travers ce manuscrit, nous cherchons à répondre à plusieurs questions qui guideront la conception d'une technologie multimodale permettant de représenter des œuvres d'art. L'objectif est d'offrir à la fois aux personnes aveugles et partiellement aveugles la possibilité de vivre leur propre expérience d'une œuvre.

— **Q1 :** Comment les informations visuelles contenues dans les œuvres d'art, telles que contours, textures et descriptions, sont-elles représentées et perçues par les personnes aveugles ou partiellement aveugles, et par quels canaux sensoriels (auditif et haptique) ces informations sont-elles transmises dans les dispositifs existants ?
— **Q2 :** Comment présenter une œuvre d'art virtuelle de manière accessible aux personnes aveugles ou partiellement aveugles ?
— **Q3 :** Comment extraire les éléments pertinents d'une œuvre d'art ?

2. Braille Authority of North America) - Guidelines and Standards for Tactile Graphics (visité le 12/03/2025)
3. Institut National Supérieur de formation et de recherche pour l'Education Inclusive) (visité le 12/03/2025)
4. INJA (Institut National des Jeunes Aveugles)-Louis Braille (visité le 12/03/2025)

3 Contributions

Cette thèse vise à développer une chaîne de traitement pour adapter des images d'une œuvre d'art pour les personnes aveugles et partiellement aveugles.

Dans ce cadre, la première contribution est un état de l'art se focalisant autour de l'interaction homme-machine pour le public visé. Cet état de l'art cherche à identifier les différents dispositifs existants dans l'exploration de documents pour les personnes aveugles et partiellement aveugles, mais aussi les méthodes de traitement spécialement conçues pour l'adaptation de différents types d'images. On étudiera notamment des représentations artistiques bidimensionnelles (e.g. peinture).

La deuxième contribution constitue des études réalisées pour évaluer les besoins et proposer un prototype de dispositif haptique 2D adapté de la tablette F2T (Force Feedback Tablet) conçue pour le projet Guide Muséal en 2019-2022 [5]. Ce dispositif haptique est une tablette à retour de force nommée F2T qui a la particularité d'être manipulée en 2D pour l'exploration des images virtuelles. Ces études ont notamment permis d'évaluer une stratégie d'exploration hiérarchique à travers ce prototype.

La troisième contribution présente une méthode d'extraction d'éléments saillants d'images de la Tapisserie de Bayeux. Pour cela un jeu de données a été collecté pour obtenir des cartes de saillance et des parcours visuels de personnes non aveugles. Cette étude permet d'étudier la possibilité de s'inspirer des stratégies d'exploration visuelles pour mettre en avant des éléments dans les images.

4 Présentation globale de la thèse

Cette thèse est composée de quatre parties principales.

La première partie (chapitre 2) présente le contexte de la littérature autour de l'accessibilité des personnes aveugles et partiellement aveugles en terme de technologie et traitement de l'information pour l'adaptation d'images.

La deuxième partie (chapitre 3) présente des études réalisées pour évaluer les besoins autour d'un dispositif haptique conçu pour favoriser l'inclusion d'un public de personnes aveugles ou partiellement aveugles dans des musées.

La troisième partie (chapitre 4) propose un jeu de données de la Tapisserie de Bayeux et son analyse pour l'adaptation à des images perceptibles via un dispositif haptique.

Enfin, la quatrième partie (chapitre 5) récapitule les contributions réalisées lors de cette thèse, ainsi que les limites et perspectives futures.

5. https ://guide-museal.univ-rouen.fr/

2 Contexte scientifique

Sommaire

1 Introduction du chapitre 16
2 Approches technologiques et sensorielles pour l'adaptation des œuvres d'art 17
2.1 Dispositifs 18
2.2 Méthodes d'exploration 26
2.3 Hiérarchisation des informations extraites 28
3 Approches pour la conception et l'adaptation de représentations d'œuvres d'art 31
3.1 Co-création de représentations et segmentation interactive 31
3.2 Chaînes de traitement automatiques 32
3.3 Les mécanismes d'attention visuelle pour l'adaptation 35
4 Synthèse 36
5 Conclusion du chapitre 41

1 Introduction du chapitre

Les œuvres d'art présentées dans les musées sont principalement destinées à être perçues visuellement. La perception visuelle peut être décrite à travers un modèle simplifié reposant sur l'intégration d'informations de bas niveau propres aux œuvres (formes, contours, textures, couleurs) et d'informations de haut niveau qu'elles véhiculent (représentations d'objets, de scènes, de symboles). Toutefois, cette perception ne dépend pas uniquement du système visuel lui-même : elle est également influencée par des processus cognitifs, culturels et contextuels qui orientent l'interprétation.

Dans ce cadre, la théorie des contingences sensorimotrices proposée par O'Regan et Noë [6, 7] s'oppose aux approches classiques fondées sur l'idée de représentations internes. Selon cette perspective, la perception n'est pas le résultat d'un traitement passif du système de perception, mais sur la manière dont les sensations changent en fonction de nos actions [8, 9].

Dans le prolongement de cette théorie, des dispositifs de substitution sensorielle (SSD - "Sensory Substitution Devices") ont été développés afin de faciliter l'accès à la perception pour les personnes aveugles ou partiellement aveugles. Le principe de ces dispositifs consiste à convertir les informations visuelles vers d'autres modalités sensorielles, telles que le toucher, l'haptique ou

l'audition [10, 11]. Toutefois, comme l'ont discuté certains auteurs [4, 5], ces SSD peinent à être globalement adoptés par le public concerné.

Plutôt que de chercher à compenser la perte de la vision par une substitution technologique, d'autres approches proposent de reconsidérer la cécité comme une source d'enrichissement perceptif. Chottin et Thompson [12] proposent, dans le cadre de l'accès à l'art, une approche différente fondée sur la notion de "gain de la cécité" ("blindness gain"). Cette perspective met de côté la notion de "perte de la vue", pour valoriser l'apport de la cécité en mettant les sens autres que la vue comme sens central dans l'expérience de l'art [13].

Comme le montrent les analyses de Uboldi *et al.* [14], bien que l'expérience des personnes aveugles face à des œuvres d'art explorables tactilement ou reposant sur la multisensorialité puisse, dans un premier temps, susciter un certain inconfort, elle conduit ensuite à une transformation de leur rapport esthétique à l'œuvre. Cette évolution rend l'expérience plus engageante sur le plan émotionnel et favorise une plus grande liberté dans le ressenti et la création d'un sens personnel.

Dans le contexte muséal, se pose ainsi la question de l'évolution des technologies pour permettre aux personnes aveugles ou partiellement aveugles d'accéder à l'art 2D. À partir de cette réflexion, nous formulons la question de recherche suivante :

Q1 : Comment les informations visuelles contenues dans les œuvres d'art, telles que contours, textures et descriptions, sont-elles représentées et perçues par les personnes aveugles ou partiellement aveugles, et par quels canaux sensoriels (auditif et haptique) ces informations sont-elles transmises dans les dispositifs existants ?

Afin de répondre à cette question, ce chapitre se structure en deux parties complémentaires et une conclusion. Dans un premier temps (section 2), nous examinerons les technologies mises à disposition à destination du public aveugle ou partiellement aveugle, leur méthode d'exploration privilégié, et le lien entre l'œuvre et la représentation par l'utilisation de différentes modalités sensorielles à travers ces dispositifs. Dans un second temps (section 3), nous analyserons l'évolution des différentes méthodes privilégiées pour l'adaptation en passant notamment par l'automatisation du processus. Dans un troisième temps, une synthèse est proposée pour les articles étudiées en présentant les limites de cette étude (section 4). Enfin, la section 5 vient clore ce chapitre en en résumant les apports principaux.

2 Approches technologiques et sensorielles pour l'adaptation des œuvres d'art

Dans cette partie, les différents dispositifs conçus pour présenter des représentations d'images d'œuvres d'art sont décrits, puis un aperçu des méthodes d'exploration selon le type de système considéré . Enfin, une section synthétise l'ensemble des dispositifs et les informations pouvant être adaptées selon leur niveau d'information.

2.1 Dispositifs

Dans cette partie, nous nous intéressons aux différents dispositifs développés pour permettre aux personnes aveugles et partiellement aveugle de vivre une expérience muséale adaptée. La section A est consacrée aux dispositifs audio, principalement accessibles aux personnes aveugles. La section B examine les représentations sonores des images. Enfin, la section C présente les dispositifs basés sur la perception haptique, qui proposent des représentations d'œuvres d'art destinées à une exploration par le toucher.

A Dispositifs audio

Dans le cadre de leur visite au musée, les personnes aveugles ou partiellement aveugles peuvent emprunter ou louer des audioguides portables contenant des audiodescriptions. Ce genre de dispositif sert à donner des éléments contextuels ou complémentaires aux œuvres présentées. Les audiodescriptions sont généralement créées par les conservateurs de musées et sont largement oculocentrées. Pour accéder à ces audioguides, les musées mettent en place des RFID (Radio Frequency Identification), ou encore des QR codes, attachés aux œuvres qui nécessitent un pointage (impossible pour les personnes aveugles).

L'utilisation d'audioguides dans les musées permet d'attirer l'attention des visiteurs sur certaines œuvres d'art [15]. Toutefois, ces dispositifs restent limités dans leur contenu : ils ne fournissent pas une description précise des objets exposés [15]. Conçus principalement pour un public non aveugle se déplaçant librement dans l'espace d'exposition, ils tendent à exclure les personnes aveugles ou partiellement aveugles. Par ailleurs, les audioguides peuvent véhiculer une interprétation particulière de l'œuvre, imposée au visiteur, au détriment de la pluralité des lectures possibles [16-18].

Dans de rares cas, certains musées proposent des audioguides fondés sur l'audio-description préenregistrée. Ce type de contenu traduit les éléments visuels (couleurs, formes, structures) en descriptions verbales accessibles aux personnes aveugles et partiellement aveugles. Selon Hutchinson et Eardley [15], contrairement aux audioguides classiques, les audio-descriptions destinées aux personnes aveugles ou partiellement aveugles offrent une information plus complète en alliant description visuelle et explication de l'œuvre. Ce format serait, de ce fait, pertinent et bénéfique pour l'ensemble des publics, y compris les non aveugles.

La pratique de l'audio-description s'appuie sur les théories de la traduction, notamment sur le principe du WYSIWYS ("What You See Is What You Say"), fondé sur une objectivité descriptive stricte [19]. Cependant, plusieurs chercheurs questionnent la pertinence de cette approche dans le contexte muséal, où l'expérience esthétique d'une œuvre dépasse largement la simple identification de ses caractéristiques visuelles (couleur, forme, taille). L'œuvre invoque des dimensions intangibles liées à l'émotion, à la cognition et à la subjectivité. Il est ainsi soutenu que la longueur, la structure et le lexique rigide de l'audiodescription traditionnelle ne permettent pas véritablement de transmettre l'expérience globale de l'œuvre [20-23].

La pratique muséale de l'audiodescription reste émergente et non régulée, néanmoins des

travaux sont réalisés dans le but d'établir une charte pour l'audiodescription d'œuvres d'art de type peinture, tapisserie et broderie en France [24]. Pour l'instant, de nombreuses interrogations persistent, tant sur les critères de sélection des contenus à décrire que sur les éléments saillants d'une œuvre à mettre en valeur et sur les modalités de leur mise en discours [25]. Malgré ces défis, des recherches ont démontré que l'audiodescription muséale peut être appréciée aussi bien par les visiteurs non aveugles que par les personnes aveugles ou partiellement aveugles [15, 26].

Les travaux de Eardley *et al.* [27] ainsi que de Chottin *et al.* [12] ont exploré l'inclusion des personnes aveugles et partiellement aveugles dans le processus de création des audio-descriptions. Cette approche remet en question le paradigme dominant WYSIWYS, souvent centré sur la perception visuelle des personnes non aveugles, au profit d'une co-construction du contenu fondée également sur la perception subjective et sensorielle des personnes aveugles.

À l'heure actuelle, les dispositifs combinant audio-description et audioguide (AD-AG) demeurent rarement utilisés dans les musées. Il convient néanmoins de souligner que des données expérimentales, tant basées sur des témoignages, suggèrent que l'expérience d'audiodescription peut être significativement enrichie lorsqu'elle est associée à une stimulation sensorielle simultanée.

Dans le cadre de visites guidées intégrant l'audiodescription en direct, de nombreux descripteurs ou guides insistent sur l'importance du toucher comme élément clé de l'expérience. Le contact tactile est perçu non seulement comme un moyen d'ancrer l'information descriptive, mais aussi comme un vecteur narratif, capable de renforcer la structure et la portée du récit. Ainsi, le toucher serait en mesure de « donner vie aux objets », tandis que des représentations haptiques simplifiées permettraient un accès initial à la structure spatiale des œuvres, que l'audiodescription viendrait ensuite compléter et enrichir [25, 28, 29].

La section suivante aborde les représentations sonores disponibles dans la littérature dans le contexte muséal.

B Représentations sonores des images dans le contexte muséal

Une grande partie des solutions d'accessibilité aux images s'appuie sur les lecteurs d'écran, largement utilisés par les personnes aveugles. L'information visuelle est ainsi transmise principalement par le biais de descriptions verbales, de sonification et, plus rarement, de retours haptiques. Ces correspondances d'une modalité à une autre correspond à la correspondance transmodale. Ces descriptions peuvent être générées manuellement [30-32], via le crowdsourcing [33-35] ou automatiquement par des algorithmes d'intelligence artificielle [36, 37]. Toutefois, les descriptions générées automatiquement restent généralement plus générales que celles fournies par des experts.

Plusieurs interfaces interactives ont été développées, comme Image Explorer par Lee *et al.*[38], ou les travaux de Kwon *et al.* [34, 35], qui permettent une navigation hiérarchique ou par attributs dans l'image. Certaines approches s'attachent à une représentation esthétique du contenu : Cho e*t al.* [39] traduisent les couleurs en poèmes et variations vocales, tandis que Jabbar *et al.* [36] utilisent un modèle d'apprentissage profond se basant sur l'architecture CLIP et des transformers pour générer automatiquement des poèmes associés à des œuvres d'art visuelles.

La sonification constitue une méthode pour transmettre des informations visuelles. Cho *et al.* [40] utilisent des mélodies pour représenter les couleurs, en fonction de leurs caractéristiques HSV, tandis que Lee *et al.* [41] explorent l'audio 3D pour coder spatialement couleur et profondeur. Toffa et Mignotte [37] proposent une approche intermédiaire combinant sons non verbaux (texture, contour) et descriptions sémantiques, mais leur dispositif reste limité aux scènes naturelles et ne prend pas en compte les œuvres d'art. Si ces approches permettent de transmettre des informations contextuelles ou descriptives, elles présentent plusieurs limites en termes d'expressivité, d'interprétation personnelle et de complexité visuelle. Par exemple, les descriptions générées automatiquement sont souvent moins précises que celles produites par des experts, avec un vocabulaire parfois inadéquat.

Les approches par crowdsourcing permettent d'enrichir les contenus mais nécessitent une main-d'œuvre importante et ne garantissent pas une qualité homogène. Des systèmes comme Image Explorer facilitent l'interaction via l'écran tactile, mais leur efficacité dépend fortement de la structuration préalable de l'image, souvent simplifiée. Les dispositifs proposant une exploration hiérarchique ou par attributs [42] ne permettent pas une navigation fine dans des œuvres visuellement complexes et reposent sur une segmentation préalable difficile à automatiser de manière fiable. D'autres travaux explorent des moyens plus créatifs de représenter des éléments visuels, comme les couleurs via des poèmes ou la tonalité sonore [39, 40] , ou encore la sonification musicale [37].

Néanmoins, ces dispositifs restent limités : leur compréhension nécessite parfois une culture musicale ou poétique, et l'absence de retour haptique prive l'utilisateur d'une perception sensorielle directe, notamment pour la reconnaissance de textures ou de formes. Enfin, les systèmes exploitant des sons spatialisés[41] ont montré une fatigue auditive et une gêne à l'usage prolongé. Ainsi, malgré leur potentiel, les dispositifs auditifs seuls peinent à restituer la richesse plastique et multisensorielle d'une œuvre d'art. L'intégration de modalités haptiques complémentaires apparaît dès lors essentielle pour dépasser ces limites et offrir une expérience plus riche et accessible aux personnes aveugles ou partiellement aveugles.

La section suivante abordera les représentations haptiques plus couramment utilisées dans le contexte muséal.

C Technologies haptiques

Cette section décrit les représentations haptiques communément utilisées par les personnes aveugles et partiellement aveugles, leurs limitations, les enjeux de conception, et le développement de dispositifs haptiques interactifs.

C.1 Représentations haptiques

Les représentations ou reproductions haptiques d'œuvres d'art sont généralement conçues par des professionnels ou des artistes spécialisés dans la production d'images adaptées à l'exploration haptique. Certains établissements muséaux proposent des ateliers spécifiquement destinés au pu-

blic aveugle, à l'instar du Museum of Modern Art (MoMA, États-Unis) [1], ou du musée Fabre à Montpellier (France), qui a notamment initié l'exposition L'Art et la matière, Prière de toucher [2].

Ce type d'exposition, centré sur l'exploration sensorielle, repose principalement sur la perception tactile des textures de divers matériaux ainsi que sur des reproductions tridimensionnelles de sculptures soigneusement sélectionnées. L'interaction avec ces dispositifs nécessite de la part des visiteurs une attention sensorielle accrue, sollicitant à la fois le toucher et l'audio-description accompagnante.

Pour rendre les œuvres accessibles au toucher, les musées emploient différents supports matériels : plastique thermoformé, papier gaufré, impression 3D, textiles ou encore bas-reliefs de peintures réalisés artisanalement. La Figure 2.1 illustre un exemple de bas-relief d'une peinture. Toutefois, les représentations haptiques d'œuvres bidimensionnelles, telles que peintures, tapisseries ou broderies, demeurent marginales dans les pratiques muséales contemporaines.



FIGURE 2.1 – Peinture Nave Nave Mahana de Paul Gauguin (a) et sa représentation en relief réalisée par Rémy Closset, AVH, France (b), présentées au Musée des Beaux-Arts de Lyon

La production de ces représentations s'avère toutefois coûteuse et/ou particulièrement chronophage [43-45]. La Figure 2.1 met également en évidence certaines limites de lisibilité tactile des bas-reliefs, notamment dans la restitution des éléments d'arrière-plan, souvent peu accessibles à l'exploration manuelle.

Une représentation haptique d'une œuvre d'art 2D devrait viser à proposer une version simplifiée, optimisée pour une lecture tactile. À cette fin, des règles spécifiques de conception ont été établies, notamment pour les supports thermogonflés. Le thermogonflage consiste à révéler un tracé en relief à partir d'un dessin noir, activé par un passage dans un four spécialisé. Les modèles peuvent être générés numériquement ou tracés manuellement à l'aide d'un feutre noir.

La Figure 2.2 présente une représentation thermogonflée d'un cheval issue de la banque d'images de l'INSEI. Elle illustre certaines recommandations en matière de transcription tactile, telle que l'espace de confort, représenté par des bandes blanches séparant les contours et les zones

1. MoMa | Accessibility (visité le 10/08/2025)
2. L'art et la matière, Prière de toucher (visité le 21/04/2025)

texturées INSEI.

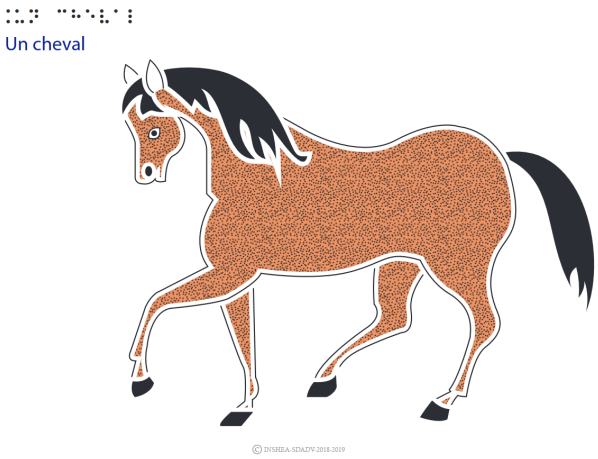


FIGURE 2.2 – Représentation thermogonflée d'un cheval. Source : INSEI

Cette technique se distingue par sa précision et sa facilité de correction. Néanmoins, elle présente certaines limites : l'impossibilité de juxtaposer plusieurs niveaux de reliefs et la fragilité du papier utilisé, qui se salit facilement.

Par ailleurs, le caractère statique, et donc non modulaire, de ces représentations limite l'accès à des détails fins : ces derniers sont difficilement perceptibles au toucher seul, en l'absence de dispositifs de visualisation dynamique, tels qu'un système de zoom. En réponse à ces contraintes, diverses techniques de présentation héritant de ces méthodes ont été développés afin d'améliorer l'accès à ces œuvres d'art bidimensionnelles.

C.2 Limites des supports haptiques et enjeux de conception

Les recherches sur l'accessibilité des images aux personnes aveugles ou partiellement aveugles se sont largement concentrées sur la production de représentations haptiques avec support imprimé ou modelé. Celles-ci prennent principalement la forme de représentations en relief, réalisées soit en bas-relief (ou 2.5D), soit via des supports thermoformés ou thermogonflés. Ces techniques permettent de traduire les éléments visuels d'une image en formes tactilement perceptibles.

Ces représentations sont produites à l'aide de procédés variés comme le fraisage CNC [17, 46], l'impression 3D [20] ou l'impression UV pour thermogonflés [47]. Cantoni *et al.* [20] ont, par exemple, réalisé des carreaux en résine imprimés en 3D représentant la tapisserie de "La bataille de Pavie", où les éléments visuels sont traduits en contours tactiles et accompagnés de légendes en braille. Cependant, cette approche purement basée sur la perception haptique, combinée à une densité élevée d'informations, peut entraîner une surcharge cognitive. Pour répondre à cette limite, Kobayashi *et al.* [47] ont proposé une représentation hiérarchisée de la gravure Sankahakuu de Hokusai, imprimée sur papier thermogonflé. Co-conçue avec des personnes aveugles, cette approche par niveaux de détail permet une lecture progressive et améliore la compréhension de l'image. Le papier thermogonflé représente en outre une alternative économique au prototypage rapide, bien que sa fragilité constitue un frein à un usage répété.

Parallèlement à la représentation de la forme, certaines recherches se sont intéressées à la conversion tactile de la couleur. Shin *et al.* [48] ont proposé une méthode consistant à convertir les caractéristiques des couleurs (teinte, saturation, luminosité) en motifs tactiles, à travers des lignes orientées, espacées et de différentes épaisseurs, fabriquées par thermoformage ou impression 3D. Néanmoins, le nombre de couleurs représentables reste limité et la correspondance avec les teintes réelles demeure approximative. Pour étendre cette gamme, Cho *et al.* [49] ont introduit des formes géométriques variées comme code de couleur, mais leur reconnaissance repose sur un apprentissage ou une légende [50], ce qui limite leur accessibilité spontanée.

Malgré leur pertinence pour l'exploration tactile, les représentations physiques en 2.5D présentent plusieurs contraintes. Elles ne sont pas modifiables une fois produites, ce qui complique leur mise à jour ou leur adaptation. Les détails fins sont parfois inaccessibles au toucher, notamment lorsqu'ils sont trop proches pour que les doigts puissent les explorer efficacement. Kobayashi *et al.* [47] ont proposé une solution en produisant des agrandissements localisés pour ces zones complexes. Enfin, les matériaux utilisés (thermoformage, thermogonflage, gaufrage, etc.) peuvent se détériorer avec le temps [1] et poser des problèmes d'hygiène dans des contextes de manipulation partagée. Ces limites soulignent la nécessité de développer des solutions interactives plus durables et adaptables, telles que les dispositifs haptiques, abordés dans la section suivante.

C.3 Dispositifs haptiques interactifs et substitutions sensorielles

La recherche sur l'accessibilité visuelle pour les personnes aveugles ou partiellement aveugles s'oriente vers le développement de dispositifs de substitution sensorielle (SSD), visant à convertir les signaux visuels en signaux perceptibles par d'autres modalités sensorielles, telles que l'audition ou le toucher [51-54].

Les approches multimodales visent à mobiliser plusieurs canaux sensoriels (e.g., tactiles, auditifs, thermiques ou olfactifs) afin de compenser la perte ou la limitation de la vision et d'enrichir l'expérience perceptive. Dans le contexte muséal, elles répondent à un double objectif : permettre aux personnes aveugles et partiellement aveugles d'accéder aux informations visuelles des œuvres, tout en favorisant une expérience immersive et émotionnelle comparable à celle des visiteurs non aveugles. Ces dispositifs s'inscrivent ainsi dans une logique d'inclusivité où la diversité des modes sensoriels devient un levier d'accessibilité.

Dans ce contexte, les dispositifs haptiques occupent une place centrale dans l'interaction multimodale, car ils permettent la transmission d'informations à travers différents types de retours sensoriels [55, 56].

Si les représentations en relief demeurent majoritairement utilisées dans les musées, plusieurs travaux ont proposé d'y associer des retours haptiques variés (vibrotactiles, thermiques ou de retour de force) afin de transmettre des informations telles que la profondeur, les textures, les contours ou les couleurs à des utilisateurs aveugles. L'un des dispositifs historiques les plus emblématiques est l'Optacon, qui convertissait des caractères imprimés en motifs vibrotactiles perceptibles par le doigt [57], bien qu'il ne soit plus commercialisé aujourd'hui.

Les dispositifs haptiques récents visent à offrir une exploration dynamique des images, grâce à

des fonctionnalités de zoom, de déplacement ou d'interaction directe avec les formes [58]. Certains s'appuient sur l'extraction automatique des contours pour générer des représentations simplifiées des éléments visuels [59-61]. Contrairement aux supports statiques, ces systèmes sont modulaires, c'est-à-dire adaptables, actualisables et capables d'évoluer selon le contenu.

La perception thermique a également été explorée comme moyen de coder certaines propriétés visuelles. Hribar et Pawluk [62] ont ainsi conçu une souris haptique intégrant une cellule braille, capable de restituer les coups de pinceau et d'indiquer les couleurs par des variations de température, bien que cette approche n'ait pas été évaluée auprès de personnes aveugles. Bartolomé *et al.* [63] ont poursuivi cette voie en utilisant des modules Peltier pour représenter la profondeur et les contrastes de couleurs à travers différents niveaux de chaleur. Leurs résultats ont toutefois mis en évidence les limites de la perception thermique, notamment le faible nombre de niveaux discernables et la difficulté à identifier les formes aux frontières des zones. Dans une approche multisensorielle, Bartolomé *et al.* [64] explorent la transmission des couleurs par mélodies et sensations thermiques. Ils soulignent cependant des limites importantes : la sensation thermique a une portée réduite, limitant la différenciation des teintes, et l'intégration simultanée des stimulations peut être difficile pour les utilisateurs. De même, Cavazos Quero *et al.* [65] combinent mélodies et parfums pour coder respectivement la teinte et la saturation des couleurs, évoquant un fort potentiel émotionnel et mnésique, mais soulignent aussi les conflits possibles liés au traitement simultané de multiples stimuli.

Les technologies d'écrans tactiles à rendu haptique ont également progressé. Certains dispositifs, comme TeslaTouch [66], Hap2U [3] et Stimtac [18, 67], utilisent l'électrovibration pour générer des stimulations vibrotactiles sur les doigts, reproduisant ainsi la sensation de textures ou de contours. Trouver le contour d'une surface indiquée par une vibration est complexe. Elle nécessite de nombreuses entrées et sorties dans la forme pour détecter les limites.

Les stimulations peuvent être modulées en fréquence selon la position du doigt, ce qui en fait des systèmes adaptatifs. Cependant, leur zone d'interaction reste limitée à un seul doigt, ce qui complique la reconnaissance des formes complexes ou la distinction entre plusieurs objets.

Plus récemment, Luo *et al.* [68] ont présenté WeSee, une interface audio-haptique modulaire combinant un affichage braille à retour vibrotactile, une surface en film TPU (thermoplastique polyuréthane) permettant l'exploration de contours gonflés par injection d'air sur une plaque imprimée en 3D, et un système narratif interactif basé sur les questions de l'utilisateur. Ce dispositif multimodal ouvre de nouvelles perspectives pour une interaction plus personnalisée et adaptative avec les œuvres d'art.

Dans le domaine de l'éducation, Petit *et al.* [69, 70] ont expérimenté des dispositifs à retour de force fonctionnant dans un espace à deux degrés de liberté (pantographe [71]), associés à des modules vibrotactiles (STRess [72]). Ces systèmes exploitent la combinaison de mouvements kinesthésiques et de stimulations vibratoires pour la lecture de graphiques ou de diagrammes simplifiés, mais restent limités à des contenus visuels déjà très épurés.

Dans le contexte muséal, l'usage des dispositifs haptiques concerne principalement les objets

3. Hap2U : https://www.vibra-nova.com/

d'art 3D, à travers des outils tels que le Phantom [73-75]. Ce dispositif permet à l'utilisateur d'explorer des images haptiques dans lesquelles des forces physiques simulées peuvent être générées à partir de modèles 3D. Il offre ainsi une restitution réaliste du relief, bien que son espace de travail demeure limité et que son usage reste restreint à des contextes expérimentaux ou spécifiques.

Des approches matérielles originales ont également été explorées. Reichinger *et al.* [76] ont proposé l'utilisation d'une matrice de spaghetti cru pour modéliser des œuvres picturales en bas-relief, mais les utilisateurs ont rapporté une gêne tactile due à la rugosité du matériau. De leur côté, Gyoshev *et al.* [77] ont expérimenté des matrices de pins de grande taille (jusqu'à 128×96) pour représenter des images tactiles en braille. Toutefois, leur résolution reste trop faible pour restituer des œuvres visuelles complexes, et leur coût de fabrication demeure élevé [**Kim_multimedia_2019**, 60, 78]. Pour améliorer la navigation et l'interprétation, Rastogi *et al.* [79] ont proposé un algorithme de zoom intuitif combiné à une souris haptique et une cellule braille, permettant d'organiser les objets d'une image selon leur proximité spatiale. Néanmoins, cette méthode reste peu adaptée aux images complexes, où la simplification des contours est difficile.

L'exploration haptique seule reste insuffisante pour comprendre pleinement une image en relief, et nécessite souvent un guide ou une description audio activable [80-83] pour être pleinement autonome. Dans cette perspective, l'intégration de nouvelles technologies électroniques permettant de créer des expériences multisensorielles engageantes apparaît comme une voie de recherche prometteuse pour aider les personnes aveugles ou partiellement aveugles à interagir avec du contenu multimédia, notamment par le biais de retours haptiques [84].

Reichinger *et al.* [82] et Buonamici *et al.* [81] proposent l'usage de caméras pour détecter les mouvements des doigts sur des modèles en 2.5D, tandis que Cavazos Quero *et al.* [83] intègrent des capteurs tactiles dans le bas-reliefs. La détection des mouvements servent notamment à activer des retours verbaux via des gestes précis.

Après avoir exploré les dispositifs haptiques interactifs et les substitutions sensorielles permettant une meilleure accessibilité des images aux personnes aveugles ou partiellement aveugles, il apparaît que l'ajout de modalités complémentaires, notamment auditives et thermiques, peut enrichir l'expérience utilisateur. Cependant, ces approches multisensorielles soulèvent de nouveaux défis liés à la complexité de l'intégration des différents stimuli et à la nécessité d'accompagner l'exploration haptique par des aides guidées.

Ces approches montrent un fort potentiel pour élargir le spectre de l'expérience sensorielle, mais soulignent également des limites importantes. Ces études mettent en avant un équilibre nécessaire entre richesse sensorielle et lisibilité perceptive : l'ajout de modalités ne garantit pas une meilleure compréhension si la charge cognitive n'est pas maîtrisée.

En somme, l'intégration de modalités sensorielles complémentaires constitue une voie prometteuse pour renforcer l'accessibilité des œuvres d'art, mais elle requiert une conception fine de l'interaction entre les sens. Dans la continuité de ces travaux, notre recherche s'intéresse à la manière dont l'association de signaux haptiques et auditifs peut être structurée pour soutenir une exploration hiérarchique, progressive et non conflictuelle des images d'œuvres, ouvrant la voie à une expérience muséale véritablement inclusive.

C'est pourquoi, il est aussi important de prendre en compte comment les personnes aveugles ou partiellement aveugles interagissent avec les différentes technologies qu'elles soient physiques ou auditives.

2.2 Méthodes d'exploration

Les supports physiques présentent l'avantage de permettre une exploration bi-manuelle, qui constitue la méthode d'exploration la plus couramment utilisée par les personnes aveugles pour appréhender les objets physiques [85]. Lederman et Klatzky [86] ont identifié sept stratégies exploratoires fondamentales employées par ces utilisateurs : le mouvement latéral, la pression, le contact statique, l'enveloppement, le suivi de contour et le poids.

Il a été démontré que la reconnaissance sémantique du contenu d'une image par voie haptique est améliorée lorsqu'elle est effectuée de manière bimanuelle [87-89]. En effet, la redondance, le parallélisme ou encore la complémentarité de l'information entre les deux mains facilitent l'intégration des données sensorielles. Cette exploration bimanuelle permet une démarche dite « organisée » : un balayage global initial conduit à une analyse plus locale, favorisant une meilleure compréhension des détails. Cette stratégie structurée a notamment été observée par Hatwell *et al.* [90] pour les enfants âgés de plus de 9 ans lors d'une étude portant sur l'identification de formes.

L'exploration locale repose essentiellement sur le gradient tactile perçu au niveau des doigts, en particulier au niveau des contours des objets. Ce gradient permet de détecter des informations fines, souvent essentielles à la reconnaissance d'un objet. La capacité à percevoir ces détails dépend directement de la densité et de la localisation des mécanorécepteurs : la densité est liée à l'acuité tactile tandis que la localisation influence la perception des irrégularités de surface, ce qui affecte directement l'orientation et l'intensité du gradient tactile.

Cependant, Orlandi [89] souligne une difficulté rencontrée par les enfants aveugles ou partiellement aveugles : l'utilisation prolongée de la stratégie de suivi de contour peut générer une accumulation d'erreurs sémantiques. En cause, la forte exigence cognitive de cette méthode, qui repose sur l'intégration d'informations ponctuelles dans une représentation globale, entraînant une fatigue plus rapide.

C'est pourquoi, au delà des stratégies d'exploration propres au support 2D, Valente *et al.* [91-93] ont réalisé plusieurs études pour tester des stratégies d'explorations liées à l'action de l'objet à percevoir, en utilisant "le gestuel des doigts pour simuler des actions" ("Action Simulations by Fingers Gestures" - ASFGs), par exemple utiliser l'index et le majeur pour simuler l'action de monter un escalier". Les auteurs [91] ont ainsi pu déterminer la pertinence symbolique des stratégies ASGFs en évaluant la capacité de personnes non aveugles à reconnaître ces stratégies réalisées par des personnes aveugles, partiellement aveugles et non aveugles aux yeux bandés. Valente *et al.* [92] ont montré que les stratégies d'exploration ASFGs sont pertinentes quels que soient les profils visuels d'utilisateurs enfants (aveugles ou partiellement aveugles) et facilitent notamment l'identification de petits objets 3D que pour des représentations tactiles utilisant des matières texturées. Les auteurs ont également évalué ces stratégies dans un livre audio-tactile augmenté avec un public d'enfants aveugles et partiellement aveugles [93]. Le livre audio-tactile "Les Doigts Qui Rèvent"

est composé de différents objets connectés avec des capteurs positionnés au niveaux des matières textiles ou des objets en 3D. Chaque objet est liée avec une action spécifique attendue selon la stratégie d'exploration ASFGs énoncée par un retour audio. Ils ont ainsi montré que les enfants quel que soit leur profil visuel ont apprécié ce support. De plus, les stratégies d'explorations attendues ont été majoritairement utilisées et ne présentent pas de différences significatives selon le profil visuel des enfants, ce qui souligne l'apport en terme d'inclusivité d'un dispositif multisensoriel.

Outre les dispositifs se basant sur des supports physiques, il existe également des dispositifs se basant sur des représentations virtuelles pour une manipulation à travers un effecteur manipulable. Plusieurs auteurs [58, 85, 94, 95] ont étudié les stratégies utilisées avec le système TACTOS. Ce dispositif, illustré dans la figure 2.3, combine deux cellules braille pour la stimulation haptique et un stylet permettant l'exploration d'une zone de travail. Il favorise diverses stratégies d'exploration, notamment des mouvements oscillatoires de différentes amplitudes.

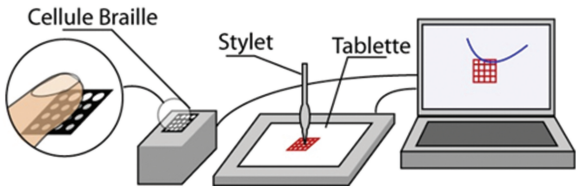


FIGURE 2.3 – Système TACTOS. Image tirée de Rovira *et al.* [96].

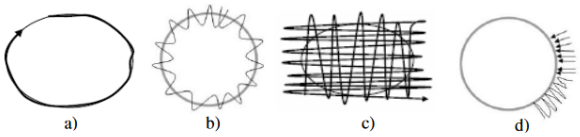


FIGURE 2.4 – Méthodes d'exploration identifiées pour le système TACTOS. a) Suivi de contours ; b) Micro-balayage, c) Macro-balayage, d) Tapotement latéral. Image tirée de Hutchinson *et al.* [95]

Ces stratégies s'inspirent directement des méthodes d'exploration manuelle précédemment décrites : "macro-balayage", "suivi de contour", "micro-balayage" et "tapotement latéral" (Fig. 2.4).

Le macro-balayage consiste en une exploration globale utilisant des mouvements amples pour repérer les principaux éléments spatiaux de la forme étudiée [9, 94, 96]. Une fois la forme localisée, les mouvements se rétrécissent progressivement, menant souvent au suivi de contour. Le suivi de contour (Fig. 2.4.a) s'effectue par un déplacement tangentiel du doigt le long des contours de l'objet, permettant d'en identifier les traits constitutifs [96]. Le micro-balayage, quant à lui, est composé de petits mouvements oscillants, généralement perpendiculaires au contour. Il a pour objectif d'amplifier le gradient tactile perçu afin d'améliorer la perception des détails fins [94]. Cette stratégie est souvent comparée aux micro-saccades visuelles, bien que celles-ci soient involontaires, tandis que les micro-balayages sont volontairement contrôlés (Fig. 2.4.c) [97]. Enfin, le tapotement latéral consiste à suivre une ligne en la touchant de façon intermittente, sans la traverser (Fig. 2.4.d).

Des stratégies similaires sont retrouvées avec d'autres dispositifs comme le TactiNet [98, 99].



FIGURE 2.5 – Méthodes d'exploration par attributs (a) et hiérarchique (b) par lecteur d'écran par Ahmetovic *et al.* [42]

Ce système a notamment été utilisé pour l'exploration de pages web, avec des retours vibrotactiles aidant à identifier la structure globale des contenus numériques [98].

Une hiérarchie similaire d'exploration est également observable dans l'utilisation des lecteurs d'écran avec interface à écran tactile [38, 42]. Par exemple, Ahmetovic *et al.* [42], en s'appuyant sur les travaux de Morris *et al.* [33], ont comparé deux méthodes d'exploration audio sur écran tactile : une approche par attributs et une approche hiérarchique (Fig. 2.5). L'approche par attributs fournit des informations ponctuelles sur un élément, tandis que l'approche hiérarchique propose plusieurs niveaux de description : contexte général, détails spécifiques, puis relations entre objets dans la scène. Cette dernière méthode a été jugée plus adaptée à la compréhension d'œuvres d'art.

Ainsi, qu'il s'agisse d'exploration physique ou virtuelle, les méthodes employées par les personnes aveugles s'appuient majoritairement sur des stratégies hiérarchisées et adaptatives, visant à optimiser la reconnaissance des formes et la compréhension sémantique.

2.3 Hiérarchisation des informations extraites

L'ensemble des travaux présentés montre que l'exploration haptique, qu'elle soit directement sur support imprimé ou assistée par un dispositif technologique, repose sur des stratégies spécifiques, souvent hiérarchisées, permettant une construction progressive de sa propre représentation. Ces stratégies, qu'elles soient bimanuelles ou unimanuelles à travers des interfaces haptiques, illustrent l'importance d'un couplage sensorimoteur efficace dans le traitement de l'information.

Le Groupe μ par le *Traité du signe visuel* [100] appuie une approche sémiotique pour laquelle un objet est perçu à travers le sens et son concept. C'est l'intégration des informations de texture, forme et couleur, qui combinées avec les informations conceptuelles fournies par la mémoire ou l'apprentissage permet de percevoir l'objet en question. Le modèle du système visuel est décrit par certains auteurs depuis longtemps comme étant la construction d'information de bas niveau vers de l'information de haut niveau par l'utilisation de divers filtres [101, 102].

Il est possible de catégoriser les modalités sensorielles utilisées pour transmettre les informations des œuvres d'art 2D.

Le tableau 2.1 synthétise les différentes modalités sensorielles impliquées dans l'extraction d'informations, en fonction de leur nature de bas niveau (formes, couleurs, textures) à haut niveau (sémantique, localisation) dans le cadre d'adaptation d'œuvres d'art 2D.

A Formes et contours

Les formes et contours des objets sont principalement accessibles par le toucher [17, 31, 32, 48, 49, 76, 82, 103, 104]. Les dispositifs utilisant des supports physiques, tels que les représentations en bas-relief, thermogonflées ou en braille, mettent particulièrement en valeur ces contours. L'audio peut également transmettre cette information, mais la perception des contours reste plus difficile [31, 32, 37, 105]. Les technologies haptiques, comme les retours vibrotactiles [62] ou de force, sont moins étudiées dans le contexte de l'adaptation artistique pour personnes aveugles, leur usage se limitant souvent à des applications éducatives avec des formes géométriques simples ou des graphiques.

B Couleurs

Les couleurs des peintures peuvent être rendues perceptibles par l'audio, via des descriptions verbales ou la sonification [31, 32, 36, 37, 39-41, 64, 65, 106]. D'autres modalités peuvent également être mobilisées : le toucher, à travers des textures [48-50], la sensation thermique [62, 64, 106], ou encore l'odorat [65]. Dans la plupart des cas, ces correspondances reposent sur une conversion de l'espace des couleurs (par exemple HSV ou RGB) vers une autre modalité sensorielle.

C Textures

Les textures peuvent être transmises par sonification [37], par retour haptique [18] ou encore tout simplement physiquement [107-109]. En effet, les textures sont souvent mis de côté par soucis de simplification pour limiter l'effet de surcharge de détails lorsqu'il s'agit d'œuvre d'art 2D telles que les tapisseries ou les peintures.

D Information sémantique et disposition spatiale

L'information sémantique des œuvres, comme le contenu ou le contexte, est généralement communiquée par des descriptions audio ou des sons d'ambiance et paysages sonores [32, 49, 65, 105]. La localisation des objets peut être rendue perceptible par l'exploration directe de supports physiques, des sons spatialisés en 3D, ou des gestes interactifs déclenchant un retour audio.

Cette diversité de modalités et de stratégies d'exploration souligne l'importance d'une conception multisensorielle adaptée, combinant intelligemment retours tactiles, auditifs et, parfois, thermiques ou olfactifs.

TABLE 2.1 – Synthèse des modalités sensorielles utilisées en fonction de l'information de l'image d'œuvre d'art 2D

Niveau	Information	Modalité sensorielle	Réferences
Bas niveau	Forme/ Contours	Audio	[31, 32, 37, 105]
		Haptique	[17, 31, 32, 48, 49, 62, 76, 82, 103-105]

	Couleurs	Audio	[31, 32, 36, 37, 39-41, 64, 65, 105, 106, 110]
		Haptique	[48-50, 62, 64, 106]
		Senteur	[65]
	Texture	Audio	[37]
		Haptique	[18, 109, 111]
Haut niveau	Information sémantique	Audio	[31, 35, 112, 113]
	Disposition spatiale	Audio	[35, 40-42, 112, 113]
		Haptique	[32, 49, 65, 105]

Les personnes aveugles se basent principalement sur leur sens du toucher et de l'ouïe pour explorer, tandis que les personnes partiellement aveugles peuvent utiliser leur vision résiduelle [114]. L'accès à l'art pour les personnes aveugles et partiellement aveugles ne se limite pas aux informations sur l'artiste et l'œuvre, mais inclut également la possibilité de rendre l'intention implicite de l'artiste accessible dans un format non visuel [39, 115]. Il est donc essentiel d'examiner les méthodes d'adaptation employées pour créer ces représentations.

3 Approches pour la conception et l'adaptation de représentations d'œuvres d'art

Cette section analyse les principales méthodes employées pour créer des représentations haptiques adaptées aux personnes aveugles ou partiellement aveugles.

3.1 Co-création de représentations et segmentation interactive

Pour les œuvres 2D, la co-création associe chercheurs, conservateurs de musées et personnes aveugles ou partiellement aveugles. Le concept de co-création est largement mis en avant en ce qui concerne la conception de dispositifs pour les personnes aveugles ou partiellement aveugles [116]. Dans le cadre d'adaptation d'œuvre d'art, Krivec *et al.* [117] suggèrent un protocole participatif dans lequel les conservateurs identifient les éléments visuels sémantiquement importants, tandis que les personnes concernées testent les prototypes et apportent des retours permettant d'en ajuster la lisibilité tactile.

Les rendus haptique générés à partir d'image s'oriente principalement vers des techniques de traitement d'images et de reconstruction 3D[118].

Reichinger *et al.* [17] ont mis au point une interface qui transforme des peintures en bas-relief. Néanmoins, le processus requiert encore des interventions humaines pour préserver le sens de l'œuvre. Ahmetovic *et al.* [42], via le système DescriVedendo, mobilisent une communauté pour décrire des images selon des règles établies par des experts. Les auteurs tracent ensuite des polygones représentant les objets (voir Fig. 2.5), puis ajoutent des descriptions verbales, produisant un espace tactile enrichi de contenu audio.

Cantoni *et al.* [20] automatisent la détection des contours via des différences de gaussiennes (DoG), mais les contours détectés nécessitent des ajustements manuels pour adapter les résultats à chaque image.

La segmentation interactive implique l'utilisateur dans l'identification des éléments visuels à rendre haptiquement explorable. Il peut dessiner manuellement des formes ou placer des repères, puis des algorithmes comme GrabCut [119] et LiveWire [120] affinent automatiquement la sélection.

Cette méthode permet de gagner du temps dans le traitement des peintures, dont les perspectives ou styles compliquent l'identification automatique de volumes. En effet, la reconstruction 3D est la méthode privilégiée pour les interfaces haptiques. Furferi *et al.* [121] combinent un algorithme de LiveWire avec des techniques de Shape From Shading pour exploiter les indices de profondeur, indiquées par les ombres, et ainsi créer des bas-reliefs. Rodrigues *et al.* [122] décrivent un processus en plusieurs étapes : segmentation interactive en utilisant GrabCut, filtrage par filtres gaussiens et de seuillage binaire pour obtenir les contours, triangulation pour le maillage, puis assignation manuelle des profondeurs.

Dans le cadre de dispositifs haptiques basés sur le retour de forces, notamment avec le dispositif

Phantom, les même procédés sont retrouvés. Li *et al.* [123] propose l'utilisation des méthodes de Shape From Shading pour générer une carte de profondeur où les textures et les contours seront rendus explorables à travers le dispositif Phantom à travers le modèle de masse-ressort de Hooke. Zhang *et al.* [124] propose des rendus haptiques d'images 2D en créant des perspectives simulées à partir du calcul des centroïdes de modèles 3D pour exploitation à travers ce même dispositif haptique. Le dispositif, à 6 degrés de liberté, traitent des forces liées à la texture, les forces de contact et de friction liés à ces modèles 3D. Des méthodes ont aussi été développées pour ne pas avoir recours à la reconstruction 3D ou des modèles géométriques.

Rasool *et al.* [125] propose d'utiliser les informations en couleur autour du point de contact avec le dispositif haptique pour calculer des forces. Les informations de couleurs sont utilisées pour obtenir une carte de profondeur. Ils traitent cette carte de couleurs en la simplifiant pour éviter des mouvements trop brusques. Pour compléter cette modélisation des rendus ils proposent d'autres méthodes pour mettre en avant les informations sémiologiques de l'image : contour/forme, couleurs et textures. Cette méthode se base principalement sur une interaction visuo-haptique.

3.2 Chaînes de traitement automatiques

Depuis les travaux fondateurs de Way et Barner [126] et Hernandez & Barner [127], la recherche vise à automatiser la conversion d'images en représentations haptiques exploitables. Ces méthodes combinent segmentation (clustering K-means, Watershed), détection de contours (Sobel, Canny), puis post-traitement adapté à la sensibilité tactile. La méthode de lignes de partages des eaux (Watershed) [127], associée à une fusion des régions, permet d'obtenir des formes fermées interprétables au toucher selon les auteurs.

Plus récemment des méthodes se basant sur des réseaux de neurones ont été étudiés pour générer des contours multi-niveaux [128, 129]. Gioia *et al.* [111] proposent de modéliser en 3D une peinture en convertissant la couleur dans une image de l'espace RGB et les coups de pinceaux en texture. La figure 2.6 présente un exemple de rendus à partir de la chaîne de traitement proposé par les auteurs. Elle se base sur la segmentation en couleurs pour extraire et convertir les textures selon les règles de création des armoires et des blasons héraldique. Un filtre de Sobel est utilisé pour détecter les contours, et ainsi garder les coups de pinceaux/traits dessinés par l'artiste. La détection de contours et le gradient de l'image permettent de générer une carte de profondeur pour la modélisation 3D. La représentation réalisée a pour objectif d'être produite sous forme de bas-relief.

Vasudevan *et al.* [130] se basent sur les contours et les textures de l'image au point de contact avec une interface ("Haptic Interface Point") pour des rendus haptiques. Le dispositif haptique à retour de force à 6 degrés de liberté utilisé est le Phantom Omni, équipé d'un stylet permettant une interaction précise avec l'environnement virtuel. Les auteurs se basent sur le contraste pour créer une image haptique, où les parties sombres indiquant les zones concaves attirent, tandis que les parties lumineuses doivent générer une force de répulsion. La figure 2.7 présente le processus de calcul des forces au point de contact. Pour obtenir les vecteurs de forces au niveau du point de contact, ils utilisent un masque de Prewitt autour du point de contact pour obtenir le gradient

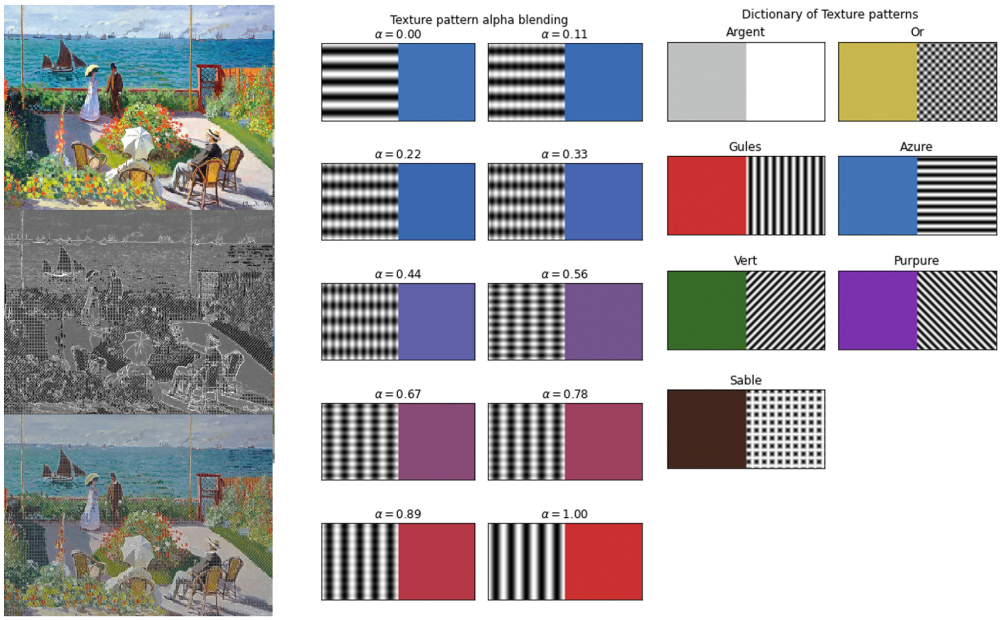


FIGURE 2.6 – Exemple de rendu des textures tirées de Gioia *et al.* [111]

dans cette zone. Ce masque est multiplié à l'image négative, et c'est les valeurs sommées des éléments qi donnent le vecteur force au point donné. Les auteurs indiquent cependant que la taille des masques a un impact important sur le bruit présent dans l'image qui impacte alors l'image haptique produite, et ainsi l'expérience des utilisateurs.

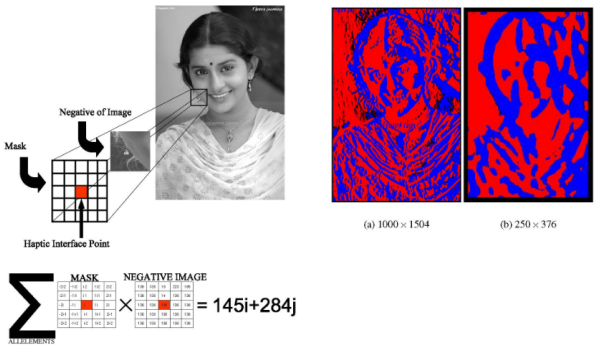


FIGURE 2.7 – Chaîne de traitement pour générer des images haptiques proposée par Vasudevan *et al.*. Image tirée de [130].

Li *et al.* [131] proposent, de manière analogue, de s'appuyer sur la loi de Hooke afin de calculer des forces à partir des composantes de couleur exprimées dans l'espace HSV. Leur dispositif expérimental repose également sur l'utilisation du dispositif Phantom Omni. Les auteurs ont évalué leur approche en testant la reconnaissance de textures constituées de stries orientées selon différentes directions, ainsi que de formes géométriques simples, auprès de participants non aveugles n'ayant toutefois pas accès à l'interface graphique. Les résultats montrent que les participants éprouvent des difficultés à discriminer les formes simples, ce qui met en évidence les limites potentielles de cette méthode pour des images plus complexes, telles que des peintures.

Kim *et al.* [132] proposent une méthode d'encodage reposant sur la manipulation des bits de poids faible (LSB) dans des images 2D, permettant d'intégrer simultanément les informations issues de l'image visuelle, d'une carte de profondeur, ainsi que de cartes de raideur et de viscosité

générées spécifiquement pour l'image considérée. Dans une étude complémentaire, les mêmes auteurs [133] appliquent cette approche au contexte des peintures interactives, en prenant pour exemple "la Nuit Étoilée" de Van Gogh. Ils génèrent alors des objets de collision aux niveaux des étoiles, en construisant ainsi un système physique associé à l'image. Leur méthode se limite toutefois aux éléments dont les propriétés physiques sont connues a priori. Dans ce cadre, les auteurs proposent d'exploiter cette chaîne de traitement à l'aide d'un dispositif à retour de force ou d'un système vibrotactile.

Les avancées en intelligence artificielle ont renforcé les approches visant à développer des chaînes de traitement permettant d'adapter les images en représentations accessibles aux personnes aveugles ou partiellement aveugles. Ferro *et al.* [43] compare des méthodes basées sur les contours comme Level Set et gPb-owt [134, 135] pour générer des représentations thermogonflées dans un cadre pédagogique. Abdusalomov *et al.* [136] combinent extraction de saillance via seuils adaptatifs à des itérations de GrabCut, améliorant la distinction des objets. Lee *et al.* [38] intègrent des modèles de détection d'objets et d'"image captioning" avec Mask R-CNN [137] et Densecap [138] pour générer des masques segmentés et légendes localisées, cependant ces solutions ne fournissent pas encore de repères spatiaux assez détaillés pour l'exploration haptique. Leo *et al.* [139] soulignent l'importance des modèles d'apprentissage profond pour la segmentation sémantique, mais pointent le manque de bases annotées spécifiques aux œuvres artistiques. Des approches de transfert d'apprentissage comme celles développées par Lu *et al.* [140] offrent une perspective prometteuse pour les adapter aux œuvres artistiques.

Dans le cas des broderies et des tapisseries où la laine utilisée donne du relief à l'œuvre, il est aussi possible d'exploiter la texture pour adapter la représentation en relief. Pizenberg *et al.* [109] ont numérisé la Tapisserie de Bayeux par photométrie stéréo, et ont produit un modèle 3D imprimable, fondé sur des annotations manuelles et des outils de maillage. Ainsi, en récupérant la texture de la Tapisserie de Bayeux, ils récupèrent une spécificité propre à l'œuvre réalisée par les brodeurs et les brodeuses. La chaîne de traitement proposée par Pizenberg est spécifique à la Tapisserie de Bayeux, et est donc difficilement applicable à d'autres types d'œuvres. Le traitement en 3D devrait s'appliquer à d'autres œuvres légèrement en relief. De plus, obtenir une représentation imprimée en relief avec un tel niveau de détail, ne permet pas une exploration haptique facile pour les personnes aveugles ou partiellement aveugles.

Redon *et al.* [107, 108, 141] a enrichi ce travail par l'automatisation de génération de cette texture en mettant en place un modèle basé sur des GAN (Generative Adversarial Network). Ainsi l'auteur propose une impression 3D sous forme de puzzle où certains éléments sont extractibles d'un socle pour faciliter l'exploration. Ainsi une proposition de chaîne de traitement est présentée intégrant la génération de texture et la segmentation des éléments de la scène en utilisant le modèle Mask R-CNN. Ici, la définition manuelle des hauteurs de chaque élément reste néanmoins nécessaire. L'arrivée de modèles comme SAM (Segment Anything Model ; [142]) permet désormais une segmentation automatique des objets dans des images efficaces, et ouvre sur des applications diverses notamment dans le contexte de l'héritage culturel pour la préservation de structures architecturaux [143, 144]. Dans le cadre des travaux de Redon, l'utilisation du modèle SAM facilite l'extraction d'objets pour la génération de modèles 3D sans besoin d'adaptation du domaine appliqué.

3.3 Les mécanismes d'attention visuelle pour l'adaptation

Le domaine de l'oculométrie permet d'étudier les schémas du regard d'une personne et les dynamiques attentionnelles. Deux grands modèles coexistent : ascendante, déclenché par les caractéristiques du stimulus (couleur, contraste), et descendante, influencé par les attentes de l'observateur.

L'utilisation de cette information pourrait servir à présenter des informations saillantes à des personnes aveugles ou partiellement aveugles. Plusieurs chercheurs ont étudié leur utilisation pour aider les personnes aveugles autour de la problématique de la navigation. Parikh *et al.* [145], Deville [146], et Yoon et Costa [147] ont utilisé la couleur et le contraste pour extraire les zones saillantes. Chakroun *et al.* [148] combinent saillance et détection de fond d'une image pour simplifier des scènes extérieures. Ces applications sont beaucoup moins fréquentes dans le cadre d'adaptation d'œuvre d'art et sont spécifiques au problématique de la navigation en temps réel des personnes aveugles.

On peut citer des exemples comme Pakenaite *et al.* [149, 150] qui proposent Pic2Tac, un système de génération d'impressions thermogonflées combinant Mask R-CNN [137], PiCANet [151] et une segmentation sémantique basée sur la base de donnée ADE20K [152]. Les objets saillants du premier plan sont extraits et associés à une librairie d'icônes, tandis que l'arrière-plan, segmenté en neuf classes, est représenté par des motifs texturaux respectant les règles pour les représentations thermogonflées.

Les expérimentations de cette étude ont été menées auprès de participants aveugles, partiellement aveugles et non aveugles. Ces derniers ont indiqué que la représentation iconographique est préférée aux simples contours, malgré des difficultés de différenciation entre objets proches et arrière-plan. Les résultats suggèrent une meilleure appréciation lorsque le premier plan et l'arrière-plan sont représentés sur des impressions séparément.

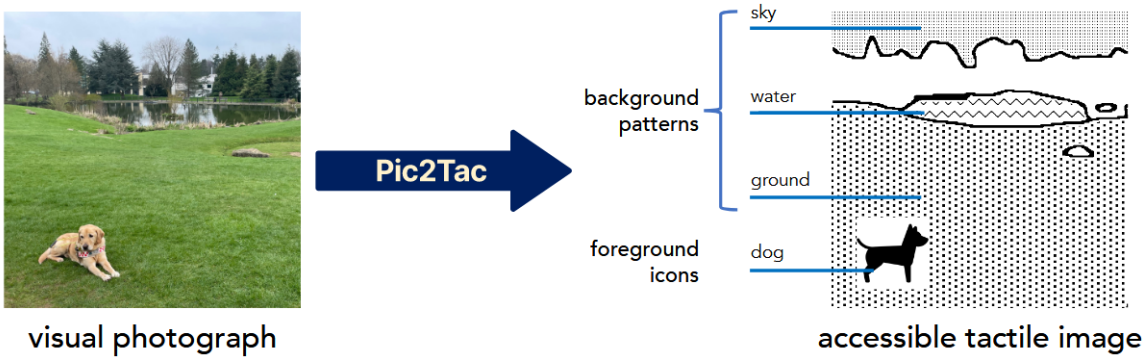


FIGURE 2.8 – Présentation de la méthode Pic2Tac avec des images tirées de Pakenaite *et al.*[150]

Les mouvements oculaires sont exploités dans des applications comme EyeDescribe [113], qui utilise les regards des visiteurs non aveugles et leur discussion pour générer des étiquettes audio spatialisées utiles pour des lecteurs d'écran. Nair *et al.* [153] proposent ImageAssist, où des cartes de chaleur issues de Grad-CAM++ [154] et ResNet18 [155] permettent d'ajuster dynamiquement les descriptions vocales selon la saillance. Dietz *et al.* [156] créent un système de sonification à

partir de données de regard.

Dans le contexte donné, les modèles de détection d'objets et de saillance (Mask R-CNN, PiCANet, Grad-CAM++) sont généralement conçus et entraînés sur des bases de données de scènes naturelles ou urbaines (ADE20K, ImageNet, etc.). Leur application directe sur des œuvres d'art peut ne pas tenir compte des spécificités artistiques qui pourraient influencer le comportement visuel comme la composition, le style, les textures etc.

De plus, il reste un manque d'intégration entre les dispositifs haptiques basés sur les retours de force et les données de regard, la plupart des solutions existantes étant adaptées à des expériences immersives destinées aux utilisateurs non aveugles. Hutchison *et al.* [157] ont utilisé des cartes de saillance basées sur l'algorithme d'Itti [158] pour moduler des vibreurs sur des sièges de cinéma pour des expériences 4D. Deng *et al.*[159] a passé en revue plusieurs types de données d'entrées pour la conception de jeux vidéos, comme les eyetrackers avec des dispositifs haptiques tels qu'une souris, un joystick, un gilet, un volant, un smartphone, un phantom avec retour de force à 3 degrés de liberté. Il constate que le développement dans ce domaine présente un grand potentiel, bien que la vision domine les dispositifs haptiques. Rantala *et al.* [160] montrent qu'un retour vibrotactile en temps réel associé au suivi du regard peut transmettre efficacement des signaux simples, mais reste limité pour les tâches complexes.

Les techniques d'adaptation actuelles présentent une grande diversité allant des approches co-construites aux traitements automatiques par intelligence artificielle. La co-création garantit une base sur des retours consistants d'utilisateurs, tandis que les segmentations interactives permettent de simplifier la chaîne de traitement pour concevoir les différentes représentations tout en gardant un savoir humain. Les chaînes de traitement automatiques ouvrent la voie à des généralisations à plus grande échelle de contenus artistiques, mais leur efficacité est encore limitée par le manque de données spécifiques au contenu artistique.

Les mécanismes d'attention visuelle offrent une base prometteuse pour la priorisation des informations à transmettre. L'intégration des données de regard avec des dispositifs haptiques ou auditifs pourrait transformer l'interaction avec les œuvres. Toutefois, les approches restent encore limitées dans le cadre de représentation d'œuvre d'art, notamment en terme de dispositifs se basant sur les retours de force.

Une convergence entre modèles perceptifs, intelligence artificielle, et retours multisensoriels semble prometteuse pour améliorer l'accès à l'image d'œuvre d'art pour les personnes aveugles et partiellement aveugles.

4 Synthèse

Le tableau 2.2 propose une synthèse des travaux portant sur les représentations d'images d'œuvres d'art en 2D. Il met en évidence les types d'informations privilégiées pour la transcription vers d'autres modalités sensorielles, ainsi que les principales techniques d'adaptation mobilisées, en distinguant celles qui sont automatisées, partiellement automatisées ou qui nécessitent une intervention humaine tout au long du processus.

Les formes et les contours constituent les informations les plus fréquemment extraites. Elles sont souvent accompagnées d'informations spatiales relatives aux relations entre les objets et à leur disposition dans l'image. Ces éléments fondamentaux sont essentiels pour des applications telles que la reconstruction 3D ou la segmentation. L'extraction des couleurs est également fréquente et permet d'être convertie en motifs texturaux pour une exploration haptique ou en retours audio. Quant à l'information sémantique, elle englobe les aspects contextuels, l'identification des objets ou des scènes, et est le plus souvent transmise via des descriptions audio. Certains travaux plus récents exploitent l'intelligence artificielle pour générer des représentations particulières, telles que des poèmes, des compositions musicales ou des interactions améliorées via des modèles de type VQA (Visual Question Answering).

Le tableau met également en avant le matériel utilisé pour supporter ces représentations et leur modularité potentielle, ainsi que leur expérimentation auprès de publics aveugles ou partiellement aveugles. Les supports thermogonflés, qui permettent une exploration bimanuelle, sont relativement peu coûteux mais se limitent à des contours simplifiés et s'usent avec le temps. Les bas-reliefs offrent davantage de détails volumétriques, mais sont souvent produits à partir d'images directement converties et présentent une modularité limitée, ce qui rend leur adaptation ultérieure difficile.

Pour pallier ces limites, certains dispositifs intègrent des capteurs tactiles ou des caméras pour suivre les mouvements des mains et des doigts des utilisateurs [32, 81, 83, 104]. Cependant, ces systèmes restent dépendants de supports non modulaires par l'utilisation de représentation bas-relief, et ne permettent pas une évolution flexible de l'expérience tactile.

Les écrans tactiles, combinés aux fonctionnalités des lecteurs d'écran, fournissent des représentations par l'audio. L'interactivité est souvent unimanuelle et limitée, mais ces dispositifs sont modulaires et intégrer dans la vie quotidienne, ce qui explique leur large adoption par les personnes aveugles ou partiellement aveugles, notamment via les smartphones.

Les dispositifs haptiques basés sur les retours de force ou vibrotactiles restent peu utilisés dans le contexte des œuvres d'art. Lorsqu'ils sont employés, il s'agit le plus souvent dans le cadre de supports pour les représentations pédagogiques simplifiées, telles que cartes ou diagrammes, utilisant des icônes et des textures pour transmettre l'information [61].

La majorité des études recensées ont testé les dispositifs auprès de publics aveugles, partiellement aveugles ou non aveugles. Certains travaux impliquent des experts aveugles ou partiellement aveugles dans la conception des dispositifs pour faire évoluer les représentations proposées [47], ou le support développé [38], d'autres proposent l'implication des artistes pour l'adaptation permettant ainsi de sélectionner des éléments pertinents en fonction du contexte ou de l'histoire de l'œuvre [82].

Cependant, la question de l'appréciation esthétique demeure peu explorée : les dispositifs permettent-ils réellement de vivre une expérience muséale riche, ou se limitent-ils à fournir des informations descriptives ? La densité et l'exhaustivité des informations transmises peuvent également restreindre la possibilité pour l'utilisateur de construire sa propre interprétation.

Cette étude se concentre exclusivement sur les œuvres d'art 2D, telles que peintures, tapisseries

ou broderies, ce qui exclut les objets pédagogiques (cartes, diagrammes, graphiques) ou l'accès à des contenus web. L'analyse de la littérature pourrait étendre dans le futur la comparaison des techniques de représentation à ces domaines.

TABLE 2.2 – Synthèse sur les représentations d'œuvres 2D telles que les peintures et les broderies pour l'accessibilité des personnes aveugles ou partiellement aveugles selon les informations ("Info.") de haut et bas niveau à extraire ("SC" = Formes/Contours, "C" = Couleurs, "Tex" = Texture, "Sem" = Information sémantique, "Sp" = Information spatiale), les modalités sensorielles principales lors de l'interaction ("Modalité") pour l'interaction (H = Haptique ; A = Audio verbal, NA = Sonification), les techniques d'adaptation (Techn. d'adaptat.) des informations extraites, si elles sont automatisées ("Auto." - X : pas d'automatisation, Δ : semi-automatisée/interactive, O : automatisée) les supports technologiques utilisées ("Dispositifs"), si des tests ont été réalisés pour avoir les retours de personnes aveugles ou partiellement aveugles ("Test"), et enfin si le système est complètement modulaire, c'est-à-dire, la possibilité de modifier la représentation sur un même support("Modulaire").

Réf.	Info.	Modalité	Techn. d'adaptat.	Auto.	Dispositifs	Test	Modulaire
Oouchi *et al.* [161] 2010	SC+Sp	H	Reconstruction 3D	X	Bas relief	Oui	Non
Reichinger *et al.*[17] 2011	SC+Sp	H	Reconstruction 3D	Δ	Bas relief	Non	Non
Hribar et Pawluk [62] 2011	SC+Sp+C	H	Correspondance transmodale	X	Souris avec cellule braille, module peltier	Non	Oui
Furferi *et al.*[121] 2014	SC+Sp	H	Reconstruction 3D	Δ	Bas-relief	Oui	Non
Volpe *et al.*[46] 2014	SC+Sp	H	Reconstruction 3D	Δ	Bas-relief	Oui	Non
Krivec *et al.* [117] 2014	SC+Sp	H	Dessin manuel	X	Thermogonflé	Oui	Non
Buonamici *et al.*[81] 2016	SC+Sem+Sp	H+A	Reconstruction 3D	Δ	Relief, camera	Oui	Non
Reichinger *et al.*[30, 76] 2016	SC+Sp	H	Reconstruction 3D	Δ	Imprimante de relief avec une matrice de 3038 Pins de spaghetti cru, caméra	Oui	Oui
Reichinger *et al.*[82, 104] 2018	SC+Sem+Sp	H+A	Reconstruction 3D	Δ	Bas Relief,Caméra de profondeur	Oui	Non
Rector *et al.*[162] 2017	C+Sem+Sp	A+NV	Segmentation, Correspondance transmodale	Δ	Camera	Oui	Oui
Cantoni *et al.* [20] 2018	SC+Sp	H	Reconstruction 3D	Δ	Carreaux/Bas-Relief	Oui	Non
Cavazos Quero *et al.*[31] 2018	SC+Sem+Sp	H+A+NV	Reconstruction 3D	Δ	Bas-relief, capteurs tactile, ventilateur, lampe	Oui	Non
Romeo *et al.*[18] 2018	Tex+Sem	H	Correspondance transmodale	X	dispositif haptique vibrotactile (stimtac)	Oui	Oui
Gyoshev *et al.*[77] 2018	SC+Sp	H	Segmentation manuelle	X	Afficheur Braille pouvant aller jusqu'à 128X96 pins	Non	Oui
Rodrigues *et al.*[122] 2018	SC+Sp	H	Reconstruction 3D	Δ	Bas-Relief	Non	Non
Iranzo Bartolome *et al.*[32] 2019	SC+Sp	H+A+NV	Reconstruction 3D	Δ	Bas-Relief, capteurs tactile, microphone	Oui	Non

Reinholt *et al.* [113] 2019	Sem+Sp	H+A	Suivi oculaire, Segmentation, Description verbale, Correspondance transmodale	Δ	Eye-tracker,écran tactile, bordure tactile	Oui	Oui
Batolome *et al.*[106] 2020	C	H	Correspondance transmodale	X	Module Peltier	Oui	Oui
Shin *et al.* [48] 2020	SC+C+Sem	H+A	Correspondance transmodale	X	Bas relief, thermogonflé	Oui	Non
Cho *et al.*[40] 2020	C	A	Correspondance transmodale	X		Oui	Oui
Cho *et al.*[39] 2021	C	A	Correspondance transmodale	X		Oui	Oui
Ahmetovic *et al.*[42] 2021	Sem+Sp	A	Segmentation manuelle	X	Écran tactile	Oui	Oui
Cho *et al.*[49] 2021	SC+C	H	Correspondance transmodale	X	thermogonflé	Oui	Non
Cavazos Quero *et al.*[83] 2021	SC+C+Sp+Sem	H+A+NV	Reconstruction 3D, Correspondance transmodale	Δ	Bas relief, capteurs tactile	Oui	Non
Bartolome *et al.*[64] 2021	C	H+NV	Correspondance transmodale	X	Module peltier	Oui	Oui
Kim *et al.* [110] 2021	Sem	A	Correspondance transmodale	O	Réseaux de neurones	Oui	Oui
Kwon *et al.*[35] 2021	Sem+Sp	A	Segmentation manuelle	X	Écran tactile	Oui	Oui
Ahmetovic *et al.*[112] 2021	Sem+Sp	H+A	Segmentation manuelle	X	Ecran tactile, lecteur d'écran, Caméra, QR code	Oui	Oui
Gioia *et al.*[111] 2021	SC+Sp+C	H	Segmentation, Correspondance transmodale	Δ	Bas-relief	Non	Non
Kobayashi *et al.*[47] 2022	SC+Sp	H	Processus manuel	X	Thermogonflé	Oui	Non
Jabbar *et al.*[36] 2022	C	A	Correspondance transmodale	O	Modèle d'apprentissage profond pour se basant sur l'architecture CLIP et des transformers	Non	Oui
Lee *et al.*[38] 2022	C+Sem+Sp	A	Segmentation d'objets	O	Ecran tactile	Oui	Oui
Redon *et al.*[107]	Sp+SC+Tex 2023	H	Reconstruction 3D, Segmentation d'objets	O	Puzzle réalisé par impression 3D	Oui	Non
Luo *et al.*[68] 2023	SC+Sem	NV+A+H	Correspondance transmodale	Δ	Actionneur pneumatique souple, vibreurs	Oui	Oui
Nair *et al.*[153] 2023	C+Sem+Sp	A	Segmentation manuelle, Spatialisation du son	Δ	Ecran tactile	Oui	Oui

5 Conclusion du chapitre

La question de recherche (Q1) ayant orienté cette revue de littérature a permis d'identifier, d'une part, les approches technologiques et modalités sensorielles mobilisées pour l'adaptation des œuvres d'art à destination des publics aveugles ou partiellement aveugles, et, d'autre part, les méthodes de conception et d'adaptation des représentations. L'analyse des travaux existants montre que le traitement des contours et des informations spatiales demeure central dans les processus d'adaptation. Ces éléments sont le plus souvent transcrits sous forme de représentations en relief, produites par impression ou gravure. Les dispositifs haptiques, quant à eux, reposent majoritairement sur une coordination visuo-tactile, ce qui limite leur pertinence pour un public privé de vision. Une approche audio-tactile apparaît ainsi plus adaptée afin de garantir l'accessibilité aux représentations.

Les méthodes d'exploration, qu'elles soient bimanuelles ou unimanuelles, soulignent l'importance d'une progression hiérarchique allant du global vers le détail. Les informations extraites peuvent être classées en deux niveaux : informations de bas niveau (formes, contours, textures) et informations de haut niveau (sémantiques et spatiales). Leur organisation, selon une double approche sémiotique et sémantique, favorise la transmodalité du sens visuel vers des modalités auditives, haptiques, et, de manière plus marginale, olfactives. Les dispositifs étudiés présentent des avantages et des limites complémentaires : les supports physiques offrent une richesse d'exploration bimanuelle mais restent peu modulables, tandis que les interfaces numériques (e.g. smartphone) sont modulaires et évolutives, mais souvent unimanuelles et limitées à la modalité auditive. L'extraction et la sélection des informations pertinentes demeurent majoritairement supervisées, nécessitant l'intervention d'une personne non aveugle afin de garantir la fidélité et la pertinence des représentations.

Les représentations haptiques, enrichies par des techniques de segmentation et soutenues par des algorithmes d'apprentissage automatique, permettent de distinguer les objets essentiels et d'accélérer la production. Toutefois, le coût de fabrication des supports demeure un frein à leur déploiement en contexte muséal. L'usage de dispositifs haptiques modulaires constitue une piste prometteuse pour proposer des catalogues de représentations diversifiés et adaptables.

Ces constats mettent en évidence la nécessité de concevoir des représentations modulaires, personnalisables et multisensorielles, intégrant à la fois des dimensions sémiotiques et sémantiques, afin de répondre à la diversité des profils des visiteurs [163]. Ils orientent également les perspectives de recherche vers le développement de solutions interactives et automatisées, capables d'assurer une expérience muséale inclusive, fidèle au sens de l'œuvre et adaptée aux contraintes cognitives des utilisateurs.

Enfin, si les avancées en vision par ordinateur et en apprentissage profond ouvrent des perspectives inédites pour renforcer l'autonomie des personnes aveugles ou partiellement aveugles, la simplification des œuvres demeure une étape incontournable. Le manque d'annotations de qualité sur les images d'œuvres d'art 2D constitue par ailleurs un défi majeur, freinant la production de représentations de haute fidélité.

3 Evaluation de l'expérience muséale

Sommaire

1 Introduction du chapitre 42
2 Étude exploratoire 1 : Définition des besoins 44
2.1 Introduction 44
2.2 Profil des participants 45
2.3 Méthode et matériel 46
2.4 Résultats 49
2.5 Conclusion 53
3 Présentation du dispositif haptique : F2T 56
3.1 Description du prototype 56
3.2 Description de l'interface 57
3.3 Définition des effets 57
3.4 Modes d'interaction 59
4 Etude exploratoire 2 : Interaction avec un dispositif haptique . 63
4.1 Introduction 63
4.2 Profil des participants 63
4.3 Méthode et matériel 64
4.4 Résultats & discussion 67
4.5 Conclusion 72
5 Conclusion du chapitre 73

1 Introduction du chapitre

Dans les années 1990, les visites guidées et les ateliers étaient les principaux dispositifs d'accessibilité envisagés par la Direction des musées de France [164]. Ces mesures restent aujourd'hui prioritaires, comme en témoigne la "Charte pour l'accueil des personnes handicapées dans les équipements culturels" publiée en 2007 par la Commission nationale "Culture Handicap". Cette charte souligne l'importance des manipulations multisensorielles, des dispositifs ludiques et pédagogiques, ainsi que des parcours interactifs dans les musées, transformant ainsi les visiteurs en acteurs. Cependant, bien que ces dispositifs permettent un accès plus autonome aux œuvres d'art, la charte omet de mentionner l'importance cruciale de l'association entre l'audio et le toucher pour une expérience véritablement inclusive [165]. Le toucher seul n'est pas suffisant pour la compréhension et l'appréciation de l'œuvre d'art : il nécessite une audio description [166].

Récemment, le concept d'accessibilité inclusive a été précédemment étudié pour rendre les œuvres d'art accessibles aux personnes en situation de handicap, et ce, dès les premières phases

de conception d'accessibilité par les travaux de Costes [167]. Toutefois, les dispositifs actuellement utilisés dans les musées restent souvent limités et peuvent stigmatiser les utilisateurs ayant un handicap car ils sont parfois les seuls à utiliser les dispositifs et surtout dans des créneaux horaires spécifiques pour les personnes en situation de handicap. L'intégration des personnes aveugles et non aveugles dans la conception de ces dispositifs pourrait non seulement en prolonger la durée d'utilisation, mais aussi en généraliser l'usage inclusif [4]. Néanmoins, il existe encore peu d'études comparant différents dispositifs pour une même œuvre d'art dans le but d'identifier les avantages et désavantages de chacun.

Les stimuli tactiles sont proposés au public lors de l'expérience du Musée de la Tapisserie de Bayeux, conjointement avec l'audiodescription, afin d'améliorer l'accès aux collections. L'objectif du Guide Muséal Inclusif est d'enrichir l'expérience des visiteurs au musée. Les visiteurs non aveugles passent en moyenne 20 minutes dans une exposition, avec des habitudes de visite superficielles : en d'autres termes, ils regardent rapidement le contenu présenté, puis s'en vont [168]. Cela correspond aux recherches montrant que le temps moyen d'observation dans une galerie d'art est de 17 secondes. Ainsi, la majorité des visiteurs non aveugles entretiennent un rapport limité et superficiel avec les musées lorsqu'aucune information interprétative supplémentaire n'est disponible. Les visiteurs aveugles ou partiellement aveugles s'appuient sur l'audiodescription. Les recherches ont également montré que l'audiodescription peut améliorer l'expérience muséale des visiteurs, qu'ils soient aveugles ou non aveugles, en renforçant le souvenir du contenu présenté [26]. Le Guide Muséal Inclusif cherche à enrichir cette expérience en intégrant aux audiodescriptions des informations tactiles, à destination des deux types de publics. Il vise à approfondir la compréhension des bienfaits de l'audiodescription, en examinant son usage combiné avec des stimuli tactiles. Lorsqu'on analyse l'impact de l'information tactile dans un contexte muséal, il est essentiel de s'interroger sur ce qu'elle cherche à accomplir. Dans le cadre du Guide Muséal Inclusif, l'objectif principal est que l'information tactile fournisse un complément interprétatif, en association avec l'audiodescription. Il est donc suggéré que les stimuli tactiles ne soient pas testés isolément.

Ce chapitre vise à répondre à la question de recherche (Q2) suivante :

— Q2 : Comment présenter une œuvre d'art de manière virtuelle et accessible aux personnes aveugles ou partiellement aveugles ?

Dans les musées, il est généralement interdit de toucher les œuvres d'art. Pour permettre l'intégration de nouvelles technologies dans ce contexte, en particulier pour s'adapter aux œuvres différentes provenant des expositions temporaires, il est essentiel de pouvoir représenter ces œuvres de manière virtuelle. De plus, cette représentation doit être accessible aux personnes présentant une incapacité visuelle, en s'appuyant sur d'autres moyens de perception.

Pour répondre à cette question, ce chapitre se concentre sur l'évaluation des besoins des personnes aveugles et partiellement aveugles en matière d'accès aux œuvres d'art.

Deux œuvres particulièrement inaccessibles car de grande taille ou derrière une vitrine, reconnues toutes les deux par l'UNESCO, ont été utilisées comme supports pour cette étude. Plusieurs études exploratoires ont été menées auprès de personnes aveugles, partiellement aveugles et non aveugles, en utilisant des dispositifs traditionnellement employés dans les musées, ainsi qu'un prototype de dispositif haptique créé dans notre laboratoire LITIS, Université de Rouen Normandie

[169].

Dans un premier temps, une étude exploratoire est présentée pour définir les besoins en supports tactiles et en retours audio. Ensuite, le dispositif haptique F2T (Force Feedback Tablet), développé pour répondre à ces besoins, est introduit. Les résultats d'une seconde étude exploratoire portant sur l'utilisation de la F2T via un mode d'interaction hiérarchique sont ensuite exposés. Enfin, une conclusion globale synthétise les résultats et ouvre la discussion sur les perspectives futures.

2 Étude exploratoire 1 : Définition des besoins

2.1 Introduction

Une étude pilote a été réalisée dans le musée de la Tapisserie de Bayeux. La Tapisserie de Bayeux, qui est en réalité une broderie, est une œuvre du XIè siècle, de grande taille, mesurant 70m de longueur pour 50cm de hauteur exposée derrière une vitre. Elle met en scène une multitude de personnages et d'objets médiévaux faisant écho à la conquête de l'Angleterre par Guillaume le Conquérant.

Une approche muséale inclusive et multisensorielle vise à rendre les musées accessibles à tous, en particulier aux personnes aveugles, en leur permettant d'expérimenter des représentations multimodales d'œuvres d'art en 2D. Cette démarche enrichit l'expérience esthétique de chaque visiteur et améliore la mémorisation des éléments observés grâce à l'exploration sensorielle approfondie des œuvres.Ces expériences sensorielles peuvent être partagées avec d'autres personnes, indépendamment de l'expérience visuelle. C'est la base même de l'inclusivité. Dans ce cadre, nous cherchons à répondre à plusieurs questions de recherche.

Les personnes aveugles ont souvent un accès partiel aux œuvres visuelles par des voies audio et tactile principalement. L'audiodescription permet de transmettre des informations verbales sur la forme de l'œuvre, sa composition et le sens de l'œuvre. Les supports tactiles, quant à eux, permettent une exploration physique. Ce qui est recherché ici est de comprendre comment la combinaison de ces deux modalités peut enrichir la perception globale d'une œuvre pour un panel de personnes avec ou sans trouble visuel. Cette problématique nous conduit à la première sous-question de recherche :

— Q2.1 : Comment l'utilisation combinée de l'audiodescription et des supports tactiles influence l'expérience d'une œuvre d'art ?

En examinant de plus près les technologies actuellement utilisées dans les musées, il est essentiel d'identifier leurs atouts et leurs limites, ceci afin d'évaluer dans quelle mesure elles contribuent réellement à faciliter l'accès à l'art. Cette réflexion motive la seconde sous-question :

— Q2.2 : Comment la technologie peut-elle être développée pour faciliter l'accès à l'art visuel pour tous ?

Il est important d'identifier des règles de conception, souvent propres au support choisi, mais qui peuvent être reprises pour la conception de nouvelles technologies. Cette démarche soulève la troisième sous-question de recherche :

— Q2.3 : Quelles seraient les règles de conception à prendre en compte pour représenter des œuvres d'art en 2D accessibles à tous ?

Pour répondre à ces sous-questions, ce travail est structuré comme suit. La Section 2.2 présente le profil des participants, en détaillant leur profil visuel et leur expérience préalable avec les dispositifs audio et tactiles. La Section 2.3 décrit la méthodologie et le matériel utilisés, incluant les supports tactiles, l'audiodescription et les technologies muséales interactives mises en œuvre pour l'étude. La Section 2.4 expose les résultats obtenus et propose une discussion sur l'influence de l'utilisation combinée de l'audio et du tactile sur l'expérience des participants. Enfin, la Section 2.5 conclut en synthétisant les principaux enseignements et en ouvrant sur des recommandations pour le développement de dispositifs muséaux inclusifs.

2.2 Profil des participants

Cette première étude a réuni 12 participants qui se sont portés volontaires, identifiés de P1 à P12. Chaque participant a signé un formulaire de consentement éclairé. Le traitement des données obtenus lors de cette étude est conforme au RGPD.

Parmi eux, 7 participants sont considérés aveugles (PA) ou partiellement aveugles (PPA) et 5 sont des personnes non aveugles (PNA). Avant de débuter les expérimentations, nous avons recueilli le profil des participants de cette étude (Tab. 3.2). Il recense, pour cette étude, la répartition des genres, le degré d'incapacité visuelle et si les participants ont précédemment visité le Musée de la Tapisserie de Bayeux. Les participants avaient une moyenne d'âge de 51.42 ± 16.64 ans.

Une catégorisation de la vue des personnes aveugles et partiellement aveugles a été réalisée. Elle est définie par l'OMS (Organisation Mondiale de la Santé) et est présentée dans le tableau 3.1. Lors de cette étudie 3 participants ont une perte complète de la vue, 2 participants ont une perte presque totale de la vue, enfin, un participant a une perte de vue moyenne et un autre une perte de vue pouvant être qualifiée de profonde. Six participants sur douze, dont un participant PA, ont déjà visité le musée de la Tapisserie de Bayeux.

TABLE 3.1 – Classification de la perception visuelle selon l'OMS

Catégorie	Type	Conditions
I	Moyenne	Acuité visuelle corrigée binoculaire <3/10 et >= 1/10 avec un champ visuel d'au moins 20°
II	Sévère	Acuité visuelle corrigée binoculaire <1/10 et >= 1/20
III	Profonde	Acuité visuelle corrigée <1/20 et >= 1/50 ou champ visuel < 10° mais > 5°
IV	Presque totale	Acuité visuelle < 1/50 mais perception lumineuse préservée ou champ visuel < 5°
V	Totale	Cécité absolue, absence de perception lumineuse

TABLE 3.2 – Tableau présentant le profil visuel des participants ; PA : personne aveugle, PPA : personne partiellement aveugle, PNA : personne non aveugle selon les catégories OMS. Genre : Masculin : M, Féminin : F. * personne ayant déjà visité le Musée de la Tapisserie de Bayeux

Participants	Profil de vision (Genre)	Participants	Profil de vision
P1	PA - cat V (M)	P7*	PA - cat V (M)
P2	PA - cat IV (M)	P8	PNA (F)
P3	PPA - cat I (F)	P9*	PNA (M)
P4*	PA - cat III (M)	P10*	PNA (M)
P5	PA - cat IV (M)	P11*	PNA (F)
P6	PA - cat V (F)	P12*	PNA (F)

Après avoir présenté le profil des participants, la section suivante décrit les méthodes et le matériel utilisés pour la collecte des données.

2.3 Méthode et matériel

Cette étude s'est déroulée dans le musée de la Tapisserie de Bayeux. Les participants ont chacun testé trois conditions d'expérience de la Tapisserie de Bayeux. Les trois conditions sont les suivantes :

— Condition A : Écoute de l'audio description (V1) en face de la tapisserie, pas de support tactile.
— Condition B : Écoute de l'audio description avec guidage tactile (V2) en face de la Tapisserie avec des supports thermogonflés
— Condition C : Écoute de l'audio description (V3) avec les hauts-reliefs.

La figure 3.1 présente les trois scènes extraites de la Tapisserie de Bayeux présentées sous les conditions énoncées ci-dessus.

Lors de cette étude, les audiodescriptions V1 et V2 de la Tapisserie de Bayeux utilisées dans les conditions A et B sont des Audio Descriptions Inclusives Cocréées (ICAD, *Inclusive Co-created AudioDescriptions*) Elles ont été élaborées par des groupes de personnes aveugles, partiellement aveugles et non aveugles [165] et s'inspirent de la théorie du "blindness gain" [**chottin___2021**, 170]. Les ICAD comprennent deux descriptions : une brève et une détaillée. La spécificité de la description brève est que la scène est décrite du point de vue d'une personne partiellement aveugle. La description détaillée est réalisée à partir des questions posées par les PA et PPA et des réponses fournies par les PNA. Ce travail collectif fait écho à celui des brodeurs et des brodeuses qui ont réalisé conjointement cette œuvre. Les descriptions détaillées suivent le sens de la narration, de gauche à droite.

Deux versions des ICAD ont été créées. Dans la version 1 (V1), les ICAD contiennent uniquement la description orale des œuvres d'art. La version 2 (V2), utilisée dans la condition B, est

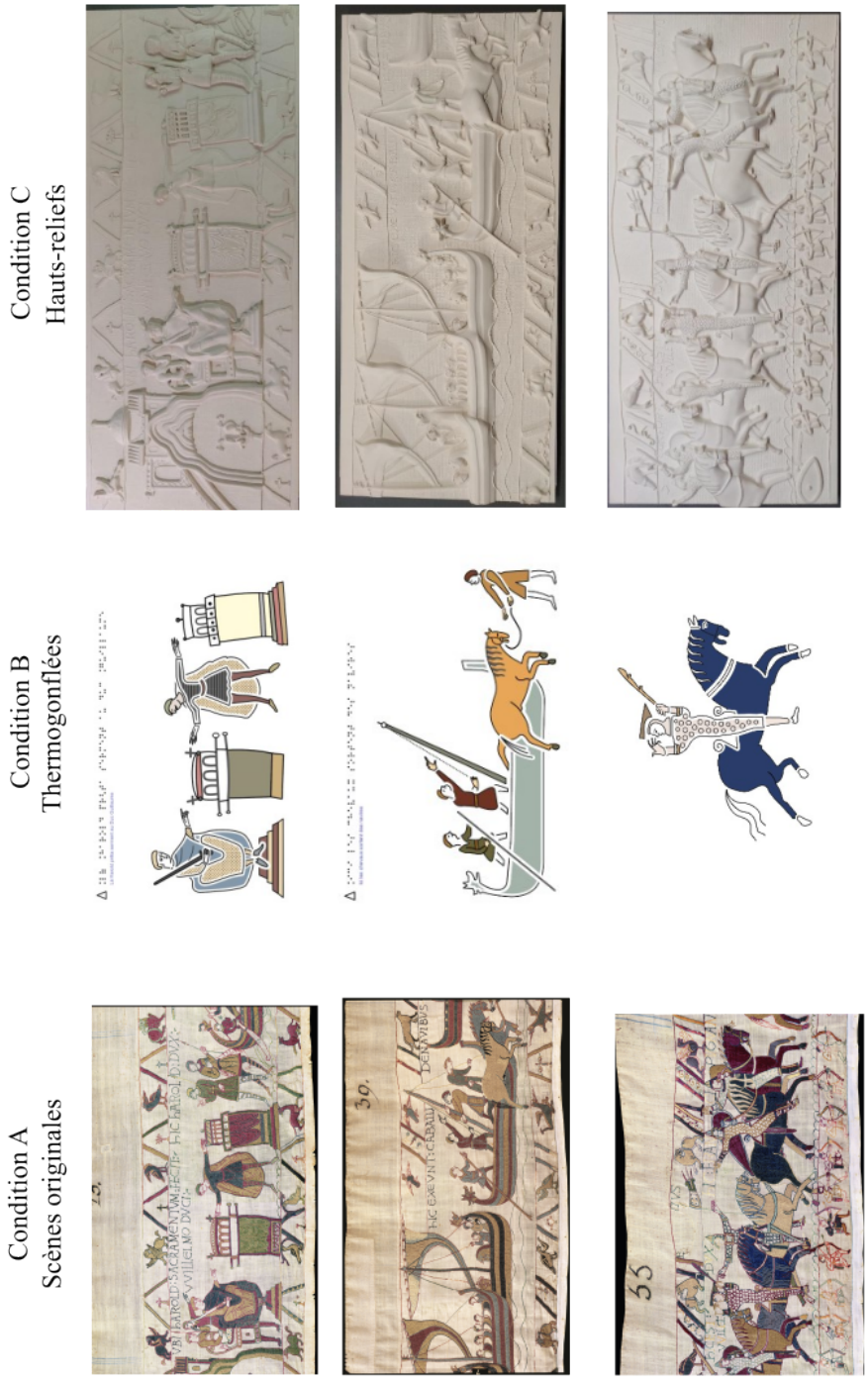


FIGURE 3.1 – Conditions expérimentales A, B et C pour les scènes 25, 39 et 55 issues de la Tapisserie de Bayeux

basée sur la V1 et incorpore un guidage de l'exploration tactile des supports thermogonflés. Ces supports en papier thermogonflé ont été conçus par le Centre Normandie Lorraine [1] et mettent l'accent sur les éléments clés de la scène de la broderie étudiée. La représentation de ces éléments est simplifiée pour permettre une discernabilité au toucher, avec des espaces de "confort" qui sé-

1. Centre Normandie Lorraine - Service Braille-Tech, https ://centrenormandielorraine.org/, dernier d'accès : 09/07/2025

parent les éléments. Pour la condition B, la version 2 (V2) des ICAD décrit la scène de manière plus générale, et intègre des indications pour l'expérience tactile.

Dans la condition C, les hauts-relief représentent toute la scène. Ces hauts-reliefs sont présents à l'étage du musée de la Tapisserie de Bayeux ; ils ont été réalisés par Rémy Closset et l'association AVH (Association Valentin-Huy) [2]. La version 1 des ICAD a été utilisée dans les conditions A et C, sans description tactile.

Après chaque condition, les participants ont rempli un questionnaire comprenant des questions à réponses fixes utilisant une échelle de Likert à 5 points pour évaluer différentes variables (incluant le plaisir, l'intérêt) (tab. 3.3) et des réponses libres permettant d'aller plus loin dans leurs expériences. Ces données ont été récoltées à partir du logiciel Qualtrics [3].

Les variables dépendantes évaluées par rapport à l'audio description sont les suivantes :

1. Plaisir (de l'audio description)
2. Intérêt (de l'audio description)

Des variables additionnelles ont été sélectionnées pour l'évaluation des stimuli tactiles :

1. Plaisir (de l'exploration tactile)
2. Intérêt (de l'exploration tactile)
3. Niveau de compréhension de l'expérience tactile
4. Amélioration de l'expérience

Les questions libres suivantes ont été posées aux participants :
— "Que pensez-vous de la manière dont le support tactile s'intègre avec l'audio description ?"
— "Aimeriez-vous être plus ou moins guidé dans votre expérience ? "
— "Que pensez-vous de la manière dont les indications pour guider le toucher s'intègrent avec l'audiodescription ?"

TABLE 3.3 – Énoncés des questions pour l'étude exploratoire 1

Questions	Énoncés
Q1	Avez-vous pris plaisir à suivre cette expérience ?
Q2	Avez-vous trouvé de l'intérêt à l'expérience ?
Q3	L'expérience était-elle compréhensible ?
Q4	L'expérience a-t-elle enrichi votre expérience de la Tapisserie de Bayeux ?
Q5	Etes-vous satisfait de la quantité d'information fournie par la scène représentée ?
Q6	Que pensez-vous du niveau de détails ?

2. Association d'aide pour les personnes partiellement aveugles et aveugles, Association Valentin Haüy, https ://www.avh.asso.fr/fr, dernier d'accès :
3. https ://www.qualtrics.com

2.4 Résultats

Les résultats obtenus suite à cette étude qualitative sont étudiés à travers l'échelle de Likert et les réponses aux questions libres posées. Ils se concentrent principalement sur l'apport des supports tactiles avec audiodescriptions fournies.

A Retours des utilisateurs

Les tableaux 3.4, 3.5, 3.6 présentent les résultats issus du questionnaire utilisant l'échelle de Likert avec une notation allant de 1 (pas du tout) à 5 (beaucoup). Les participants ont généralement montré une grande satisfaction pour toutes les conditions, mais quelques différences peuvent être observées. On se basera alors principalement sur l'étendue des données.

TABLE 3.4 – Résultats issus de l'échelle de Likert pour les énoncés Q1 à Q3 présentés dans le tableau 3.3 (notation : 1 - "pas du tout" ; 2 - "un peu" ; 3 - "neutre" ; 4 - "assez" ; 5 - "beaucoup" ;

Conditions	Condition A médiane [min,max]	Condition B médiane [min,max]	Condition C médiane [min,max]
Q1 : Avez-vous pris plaisir à suivre cette expérience ?	5 [2,5]	5 [4,5]	5 [2,5]
Q2 : Avez-vous trouvé de l'intérêt à l'expérience ?	5 [4,5]	5 [4,5]	5 [2,5]
Q3 : L'expérience était-elle compréhensible ?	5 [2,5]	5 [2,5]	4 [2,5]

Pour la question Q1 sur le plaisir ressenti, la valeur minimale est plus basse pour les conditions A et C (min =2), comparée à la condition B (min = 4).

En explorant le niveau d'intérêt des participants (Q2), les scores sont tous très proches, suggérant un haut niveau d'intérêt pour toutes les conditions (med = 5). Ici, les conditions A et B ont la même valeur minimale (min =4) et sont légèrement supérieures à la condition C (min=2).

Toutes les expériences ont été jugées très compréhensibles (Q3). Les conditions A et B ont une médiane de 5 [2, 5], avec une très légère réduction de la compréhension pour la condition C (med=4 [2,5]).

TABLE 3.5 – Résultats issus de l'échelle de Likert pour les énoncés Q4 et Q5 présentés dans le tableau 3.3 (notation : 1 - "pas du tout" ; 2 - "un peu" ; 3 - "neutre" ; 4 - "assez" ; 5 - "beaucoup"

Conditions	Condition B médiane [min,max]	Condition C médiane [min,max]
Q4 : L'expérience a-t-elle enrichi votre expérience de la Tapisserie de Bayeux ?	5 [2,5]	5 [3,5]
Q5 : Etes-vous satisfait de la quantité d'information fournie par la scène représentée ?	5 [3,5]	4 [2,5]

Les conditions B et C proposent en plus d'une expérience tactile l'audiodescription présentée dans la condition A. Ainsi, les questions Q4 et Q5 cherchent à évaluer cette modalité supplémen-

taire. En terme des avis donnés sur les conditions B et C qui ont semblé enrichir leur expérience de la Tapisserie de Bayeux (Q4), les avis sont positifs (med = 5) pour les deux conditions citées avec une variabilité plus importante. En effet, la condition B a une valeur minimale (min = 2) légèrement plus basse que la condition C (min = 3).

C'est la condition B qui semble avoir satisfait le plus les participants pour la quantité d'information fournie (Q5) comparée à la condition C.

TABLE 3.6 – Résultats issus de l'échelle de Likert pour l'énoncé Q6 présenté dans le tableau 3.3 (notation Q6 : 1 - "pas assez de détail"s ; 2 - "assez peu de détails" ; 3 - "assez de détails" ; 4 - "un peu trop de détails" ; 5 - "trop de détails")

Conditions	Condition A médiane [min,max]	Condition B médiane [min,max]	Condition C médiane [min,max]
Q6 : Que pensez-vous du niveau de détails ?	3 [3,4]	3 [2,5]	3 [2,5]

Comme indiqué dans le tableau 3.6, à la question Q6, les participants ont trouvé que les représentations avaient suffisamment de détails pour toutes les conditions, la condition A ayant la plus petite plage ("assez de détails" med=3 [3, 4]). Les audio descriptions (condition A) sont considérées comme ayant la bonne longueur. Le haut-relief (condition C) présente certains détails qui ne sont pas toujours indiqués par la description audio et qui ont été signalés par plusieurs participants, ceci peut être relié à la quantité importante d'information tactile sur ce support.

B Analyse des besoins

Les réponses des utilisateurs face à des questions libres permettent de mieux nuancer les scores élevés donnés précédemment. Ce paragraphe regroupe les divers commentaires émis par les participants. Ils répondent aux problématiques suivantes :

— Comment l'utilisation combinée de l'audiodescription et l'exploration tactile influent sur l'expérience ?

— Quelles sont les règles pratiques pour la réalisation de représentations tactiles et haptiques ?

B.1 Influence des dispositifs audio-tactile sur l'expérience muséale

Certains participants aveugles et partiellement aveugles ont relevé qu'il n'était pas utile de venir au musée pour uniquement écouter une audiodescription qui pourrait être mise en ligne, un support tactile peut donc apporter une valeur ajoutée à l'expérience muséale des visiteurs. D'un point de vue pratique, les participants ont indiqué une préférence pour la Condition B, où l'expérience tactile était intégrée avec l'expérience audio issue de la description ICAD et un guidage tactile réalisée spécifiquement pour cette condition. Contrairement à la condition C qui a été présentée avec la même audiodescription ICAD mais sans guidage tactile additionnel.

P1 : *" [..]Le thermogonflé est un support pratique."* ;
P3 : *"Le thermogonflé permet aux partiellement aveugles de mieux comprendre visuellement ensuite tactilement."*

Mais cela implique aussi une possible différence entre ce qui est représenté et ce qui est écouté. Cela peut être expliqué par le type de représentation choisi, notamment les supports thermogonflés qui nécessitent d'extraire un nombre limité d'éléments pour faciliter la compréhension par le tactile :

> P2 : *"Deux personnages mentionnés dans l'audio mais pas présents sur le dessin d'où la lecture tactile perturbée. Peut-être faudrait-il les ajouter sur le dessin ou indiquer leur absence dans l'audio description."* ;
> P7 : *"Il est perturbant d'avoir des éléments décrits qui n'apparaissent pas."*.

Pareillement pour la condition C, l'audio description apporte une meilleure compréhension de la scène même si l'audio description ne décrit pas précisément ce qui est touché tactilement. Néanmoins, l'absence de couleurs sur le haut relief peut porter à confusion, puisque l'audio description les décrit. Ce point revient souvent :

> P8 : *"Il y a un problème lié à la référence aux couleurs."* ;
> P10 : *"Il y a des couleurs qui n'apparaissent pas, il faudrait les enlever des audio descriptions."*

La mention de la couleur ne serait pas obligatoire pour certains, et peut même perturber les personnes qui voient :

> P1 : *"[...]Les couleurs ne sont pas représentées : la précision des couleurs n'est pas obligatoire à l'oral. Le choix de présenter les couleurs devrait être selon la personne ou son profil (visuel)."*.
> P3 : *" J'apprécie beaucoup l'association audio et relief. Pour un voyant, le descriptif des couleurs perturbe."*.

À la question de savoir quelle condition ils préféraient, trois participants (P1, P4, P10) ont mentionné que l'audio description seule (condition A) suffirait mais, le participant P1 ajoute qu'un modèle tactile serait un plus, car il rendrait certains éléments représentés tactilement plus proéminents. Les huit autres participants ont préféré la combinaison de la description audio et d'un support tactile. En effet, comme noté précédemment, la multimodalité (audio et tactile) est considérée comme une expérience enrichissante (Condition B (med= 5 [3,5]) et Condition C (med= 5 [2,5])).

Parmi les huit participants, plusieurs participants (P2, P6, P9) ont eu du mal à choisir un unique support tactile, car ils estiment que leur usage n'avait pas la même finalité : la condition B a plus une voie pratique tandis que la condition C a plus un usage esthétique. Cet aspect esthétique est notamment relevé par des participants non aveugles (P11 et P12) :

> P11 : *"En tant que voyant, j'ai une préférence pour le haut relief qu'on peut toucher et qui donne une idée des volumes."*
> P12 : *"Je préfère l'audio description avec le haut relief parce que ça réhausse ce qui est important dans la scène."*

B.2 Intégration du tactile et de l'audiodescription

Les questions suivantes sur l'intégration du support tactile avec l'audio ont été posées aux participants dans le cadre de la condition B et la condition C.

En regardant les retours des participants par rapport à la question libre : *"Que pensez-vous*

de la manière dont le support tactile s'intègre avec l'audio description ?" posée dans le cadre de la condition C, l'audio description est jugée indispensable pour les supports tactiles, notamment pour le haut-relief où les éléments ne sont pas simplifiés et comportent le plus de détails des scènes de la Tapisserie de Bayeux :

> P1 :*"L'audio description devrait être en complément. Les deux supports (tactile, audio description) devraient être explorés en même temps (Condition C). Une description détaillée aide à la compréhension et à se rappeler des éléments."* ;

Mais même dans ce cas, il est difficile d'appréhender les haut-relief en raison de la quantité d'éléments présents. Toujours pour la condition C, les participants ont commenté ceci :

> P2 : *"Il manque de précision quant à la position des éléments. l'audio description ne mentionne pas les bandes hautes et basses (de la Tapisserie)."* ;
> P8 : *"Il y a beaucoup d'éléments un peu tassés, pas assez séparés"* ;
> P12 : *"Il y a trop de choses inutiles dans l'audio : parler des moustaches est inutile"*.

La remarque du Participant P12, nous indique que l'information liée au contexte n'a pas été communiquée. Les moustaches distinguent les anglais des normands dans la tapisserie, nous avons considéré ce détail comme important.

On peut en conclure que pour chaque élément représenté, il est nécessaire de donner les explications et de guider davantage l'exploration tactile.

> P3 : *« Cela (audiodescription) apporte un plus. Le relief seul serait compliqué à comprendre »* ;
> P4 : *« Sans audiodescription, le haut relief est illisible »* ;

La synchronisation entre l'exploration tactile et l'audio description avec le guidage tactile permettrait de mieux situer l'information sur le support et de savoir ce que représente l'élément touché.

P2 : *« Il manque de précision quant à la position des éléments »*.

A la question libre :*"Aimeriez-vous être plus ou moins guidé dans votre expérience ?"*, dix participants sur douze ont exprimé vouloir être plus guidé. Et ils ont émis les remarques suivantes, pour la condition C :

> P1 : *« J'aimerais être un peu plus guidé pour localiser l'endroit où poser les mains. »*
> P2 : *« Il faudrait être plus guidé, plus précisément que présentement, avec des indications de direction plus précises. »*
> P3 : *« Cela me permettrait d'avoir une base (vue globale de la scène). »*
> P7 : *« J'aimerais pouvoir vérifier ce que je touche. »*

Les retours montrent que les participants souhaiteraient être plus guidés dans leur exploration tactile du haut-relief, notamment pour avoir des indications plus précises sur ce qui est touché, et ainsi avoir une vue globale et détaillée de la scène.

Lors de l'introduction du guidage tactile dans l'audio description de la condition B, la question libre suivante a été posée aux participants : *Que pensez-vous de la manière dont les indications pour guider le toucher s'intègrent avec l'audiodescription ?* Dix participants sur douze ont apprécié ces éléments de guidage, les deux autres (P4, P9) ont jugé que l'audio description était trop rapide, et ne laissait pas assez de temps pour explorer.

P3 : *« Cela permet un guidage de base. L'audio permet vraiment de ne pas se perdre dans les détails. En tant que personne partiellement aveugle, le schéma simplifié (thermogonflé) aide beaucoup. ».* ;
P5 : *« [...] Il aurait été bien d'explorer à partir d'en haut à gauche, par Guillaume, pour passer d'un objet à un autre. »*
P7 : *« Tout est bien sauf la description des reliquaires, on est au niveau des mains de Harold ce serait mieux de commencer leurs descriptions par le haut ».* ;

Effectivement, des interrogations se sont posées suites au choix du cheminement de l'audio description de la scène 23 qui commençait par le bas car les personnages auraient pu être repérés plus facilement par les pieds au toucher. Par rapport à la disposition de la feuille, il nous semblait qu'il était plus facile de commencer par le bas.

Il faut retenir ici que l'audio description devrait avoir un cheminement logique dans le sens de l'écriture de gauche à droite, de haut en bas.

B.3 Elements de représentation

L'une des problématiques lorsqu'il s'agit de réaliser des représentations est de savoir comment présenter les informations dans l'image. Il est possible de remarquer que la simplicité des représentations thermogonflées, qui se base principalement sur des contours mis en relief et des textures utilisées appliquées en suivant les recommandations pour la transcription de documents (Source : INSEI), est appréciée d'un point de vue pratique d'après ce qui a été remarqué précédemment.

(condition B, scène 23) P5 : *"[...]La texture pointillée de l'habit est très bien. Les marches sont bien marquées. Le personnage de gauche est plus haut et bien représenté (par sa hauteur)."*)
(condition B, scène 55) P9 : *"Bonne texture qui permet de mieux percevoir et les contours sont plus faciles à suivre."*)

Les contours servent principalement à guider l'exploration tactile en suivant l'audio présentant un guidage tactile et des éléments détaillés. Le relief utilisé dans la condition C est apprécié car il permet de fournir une information supplémentaire sur le volume des éléments représentés et donc rehausser ces derniers pour les rendre plus saillants.

P1 : *"Les nombreuses textures sont pertinentes et aident à rendre l'expérience tactile compréhensible. Le relief a l'avantage d'avoir différents niveaux d'épaisseur et la matière est agréable."*
P11 :*"Le support tactile (condition C) apporte vraiment un plus à la compréhension des volumes et des formes."*

2.5 Conclusion

Les différentes conditions de test sont généralement bien appréciées par les participants comme il est possible de voir à travers les retours des participants. Tandis que les réponses libres permettent de dégager des recommandations intéressantes pour les méthodes d'interaction à prendre en compte pour développer un dispositif haptique. Suite à ces retours, il est possible alors dégager des réponses possibles aux questions de recherches posées au début de cette section.

— Q2.1 : Comment l'utilisation combinée de l'audiodescription et des supports tactiles affecte l'expérience de la Tapisserie de Bayeux ?

L'étude qualitative réalisée ici a permis de mettre en avant l'influence positive sur l'expérience muséale des représentations multimodales (utilisant des supports tactiles et audio) dans le contexte du Musée de la Tapisserie de Bayeux. L'intégration d'indications tactiles dans le texte de l'audio description pour les représentations thermogonflées permet d'avoir un support pratique tactile permettant de se concentrer sur la scène décrite. De même, les présentations en haut-relief, qui tendent vers un support plus esthétique, apportent aux personnes qui l'utilisent avec une audio description, une première explication plus générale décrivant la scène. La majorité des participants, qui ont déjà visité la Tapisserie de Bayeux ou non, ont trouvé enrichissant les conditions expérimentales présentées avec une préférence vers les conditions C et B qui apportent une plus-value tactile à l'expérience audio du visiteur.

— Q2.2 : Comment la technologie peut-elle être développée pour faciliter l'accès à l'art visuel pour tous ?

Les technologies devront proposer une exploration tactile guidée suivant les audio descriptions. Pour guider l'exploration, il serait nécessaire de synchroniser l'audio description à l'exploration tactile. Les contours seront utiles à exploiter pour qu'une personne ait une idée des formes représentées, et/ou explore en détail ces éléments. Une personnalisation est nécessaire selon le profil des participants, notamment en termes de description détaillée (e.g. description des couleurs). Une personne en explorant tactilement devrait avoir accès à plusieurs niveaux de détails (audio description brève ou détaillée).

Pour cela, la description devra porter sur l'ensemble de la scène présentée. Les éléments saillants devraient être facilement discernables par leur textures ou des effets de reliefs, comme indiqué par les règles établies pour les supports tactiles [4] [5]. Une version simplifiée devrait être disponible, néanmoins, la scène devrait aussi être présentée au complet [33, 42]. Pour cela, il serait intéressant d'avancer le concept de niveau de "gist", l'information essentielle de l'élément ou d'une scène. Le premier niveau de gist sert à présenter une version simplifiée de la scène avec les éléments saillants présentés avec une description essentielle et rapide. Le second niveau de gist regrouperait les détails essentiels et nécessaires pour une représentation plus complète de la scène. Ceci rejoint les recherches menées autour de l'accessibilité des images à des PA. Le choix du type de support : audio et tactile ou audio seulement devrait être disponible pour tous. Il faudrait donc pouvoir proposer des interactions permettant une exploration libre, mais avec un apport d'information sur ce qui est touché, et une exploration guidée permettant de situer les éléments entre eux [171].

— Q2.3 : Quelles seraient les règles de conception des représentations d'œuvres d'art 2D accessibles à tous ?

Pour la création de représentations multimodales, il est nécessaire de suivre certaines règles de conception. Ci-dessous une liste des recommandations non exhaustive pouvant être relevée d'après les commentaires émis par les participants :

4. insei.fr (visité le 12/03/2025)
5. Braille Authoritity : Guidelines and Standards for Tactile Graphics (visité le 12/03/2025)

- Utiliser le relief ou les textures pour marquer les éléments importants ;
- Utiliser les contours pour suivre tactilement les contours des formes pendant l'audio description ;
- Annoncer les éléments représentés et ceux qui ne sont pas représentés ;
- Permettre la sélection d'éléments pour une exploration détaillée approfondie ;
- Synchroniser l'audio et le tactile ;
- Permettre d'obtenir des informations de localisation sur les formes présentées ;
- Proposer un mouvement logique dans l'exploration d'une scène.

Cette étude a permis de fournir des informations contextuelles essentielles pour mieux intégrer les personnes aveugles et partiellement aveugles (PA) dans leur expérience muséale, en même temps que les personnes qui n'ont pas d'incapacité visuelle. Une évaluation pilote a été menée au musée de la Tapisserie de Bayeux, où douze participants ont exploré trois scènes de la Tapisserie à travers trois conditions différentes, utilisant des audio descriptions inclusives et co-créées (ICAD) ainsi que des représentations tactiles (papier thermogonflé et haut-relief).

Les résultats ont montré que chaque condition était perçue comme intéressante par les participants. Cependant, plusieurs améliorations sont nécessaires, en particulier, concernant la conception des présentations multisensorielles.

Dans le but de préparer les futurs tests avec le dispositif haptique F2T, cette évaluation exploratoire a été menée dans un environnement muséal, sous différentes conditions de test, afin de recueillir les retours d'un public incluant des personnes aveugles, partiellement aveugles et non aveugles.

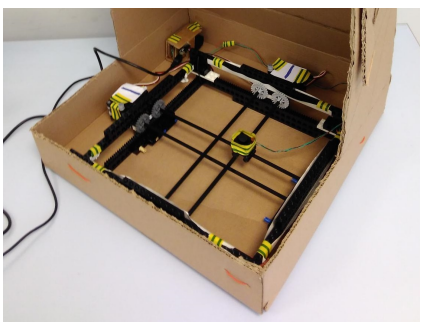
L'objectif de cette étude exploratoire était de comprendre comment les personnes aveugles ou partiellement aveugles, ainsi que les personnes qui voient, perçoivent les mêmes descriptions audio et représentations tactiles dans un contexte muséal. Cette recherche apporte des renseignements précieux sur la manière dont différents types d'interprétations multisensorielles peuvent être optimisés pour améliorer l'accessibilité des œuvres d'art dans les musées. Néanmoins, à cause du nombre de participants limité, il n'est pas possible de généraliser les recommandations obtenues.

3 Présentation du dispositif haptique : F2T

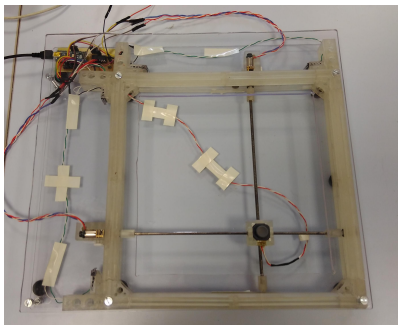
Cette section présente la tablette haptique F2T (Force Feedback Tablet), son fonctionnement et les ressentis haptiques générés. La F2T (Force Feedback Tablet), une tablette haptique à retour de force, a été développée pour induire la perception des formes via un retour de force appliqué par un joystick, accompagné d'un retour audio guidant l'exploration [169]. Une proposition d'une méthode d'interaction est explicitée à partir de ses fonctionnalités.

3.1 Description du prototype

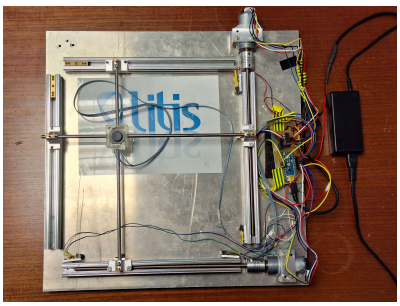
La F2T (Force Feedback Tablet) est un dispositif haptique développé par Gay *et al.* [169, 172]. La première version de ce dispositif a été réalisée à partir de briques de Lego (Fig. 3.2a), suite à cela la deuxième version utilise l'impression 3D en résine pour ses pièces (Fig. 3.2b). Dus à des soucis de solidité et de conception, une version métallique du prototype F2T a été développée (Fig. 3.2d). Celle-ci comporte un cadre métallique formé de rails de 22×22 cm. Deux de ces rails sont reliés à des moteurs à courant continu qui actionnent des vis sans fin, mettant en mouvement des composants pouvant glisser sur les rails et déplacer le joystick. Ce joystick à tête plate se déplace sur un plan 2D .



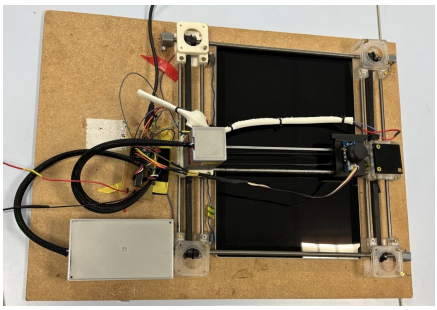
(a) Version Lego



(b) Version résine



(c) Version métallique



(d) Version courroies

FIGURE 3.2 – Prototypes de la F2T

Un moteur actionne le mouvement du joystick sur l'axe X et l'autre sur l'axe Y. Aux extrémités des deux axes, des capteurs de fin de course ont été placés pour fonctionner dans le cadre où l'image virtuelle est représentée haptiquement. Toute l'électronique est contrôlée par une Arduino Nano. Le pilote qui actionne les moteurs est alimenté par une alimentation de $9V$.

3.2 Description de l'interface

Ce dispositif est connecté à une interface écrite en Java qui traite des images haptiques virtuelles (Fig. 3.3). Certaines régions de l'image sont identifiées comme des *"zones d'intérêt"* qui sont renseignées manuellement, fournissant un retour haptique lorsqu'elles sont explorées. La F2T fournit des illusions de friction et de hauteur par le biais des retours kinesthésiques. Ces effets sont gérés par une classe dédiée aux images virtuelles, chaque canal RGB des pixels représentant des propriétés distinctes. Des classes supplémentaires gèrent des effets tels que les flux et les rails, encodant des champs vectoriels via les canaux rouge et vert de chaque pixel.

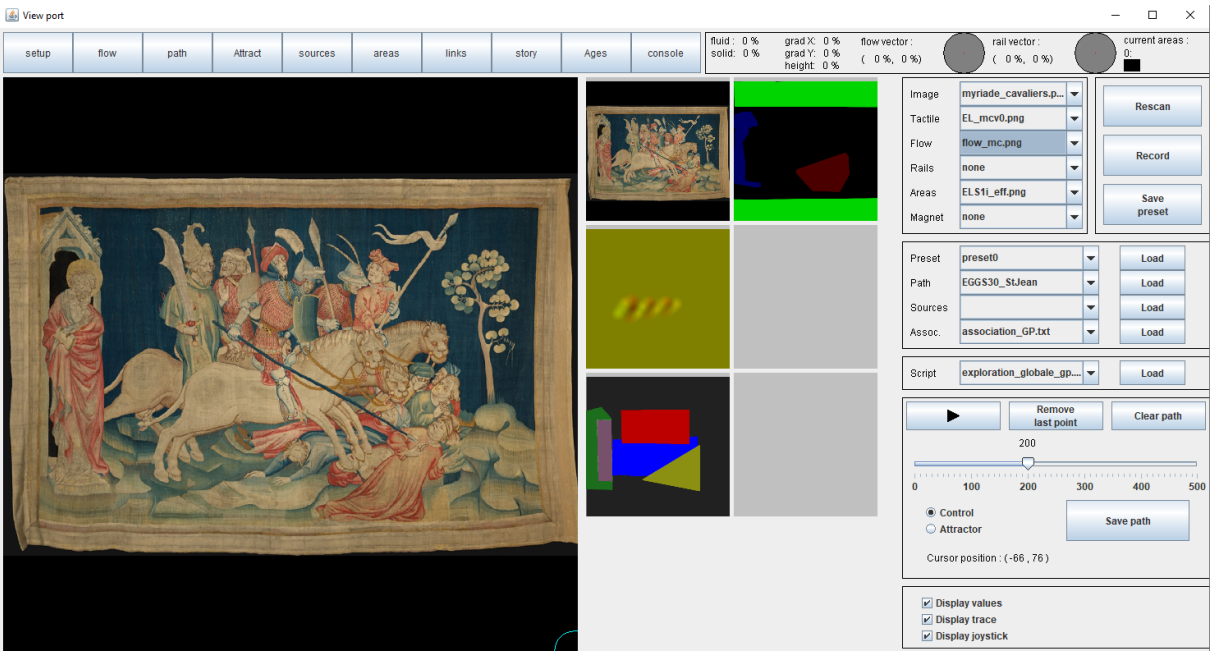


FIGURE 3.3 – Interface de la F2T

3.3 Définition des effets

Les retours haptiques sont créés selon le mouvement du joystick engendré par l'utilisateur (cf. Fig. 3.4). La différence en position $\vec{J}$ est calculée au niveau des potentiomètres du joystick entre

le support et le capuchon du joystick : $\vec{J} = P_{doigt} - P_{support}$. C'est cette grandeur qui permet de commander en vitesse les moteurs à travers une commande de type PID.

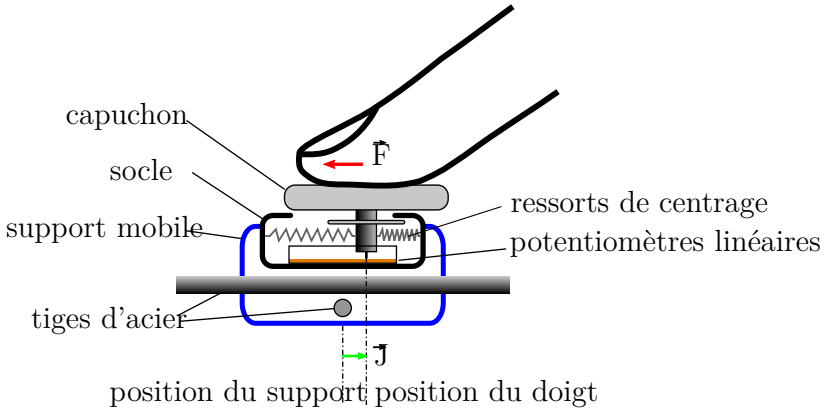


FIGURE 3.4 – Système du joystick permettant de récupérer le mouvement engendré par le doigt [169]

La vitesse $V(\vec{x}, y)$ des moteurs commandés est définie de la manière suivante :

$$\vec{V} = \alpha.\vec{J} \tag{3.1}$$

où α est la vitesse maximale des moteurs.

A Effets de friction

Deux types d'effets frictions sont proposées et se basent sur la réduction de la vitesse du support. Les équations 3.2 et 3.3 régissent, respectivement, le comportement pour l'effet de friction fluide, simulant la viscosité, et l'effet friction solide, simulant les surfaces rugueuses. Les coefficients de frictions fluide $f_f(B(x_p, y_p)))$ et solide $f_s(R(x_p, y_p))$ varient dans l'intervalle $[0; 1[$, avec $R(x_p, y_p)$ et $B(x_p, y_p)$ qui représentent les valeurs des canaux rouge et bleu du pixel de l'image situé à la position du joystick.

Dans la partie logicielle, ces coefficients sont renseignés d'après des valeurs des pixels selon les canaux bleu et rouge dans l'image virtuelle pour l'effet de friction solide (Eq. 3.3) et fluide (Eq. 3.2) respectivement.

$$\vec{V}' = \vec{V}(1 - f_f(x, y)) \tag{3.2}$$

$$\vec{V}' = \begin{cases} \vec{V}(1 - f_s.\frac{1}{||\vec{V}||}), & \text{si lorsque } ||\vec{V}|| > f_s \\ 0, & \text{sinon} \end{cases} \tag{3.3}$$

où $\vec{V}$ correspond à la vitesse du joystick manipulé par l'utilisateur, et $\vec{V}'$ correspond à la vitesse du joystick modifié à travers les effets haptiques appliqués.

B Effet de hauteur/mur

L'effet de hauteur/sensation de bosse est un effet connu se basant sur la variation de la vitesse qui dépend d'un gradient d'une hauteur h (Eq. 3.4). Ce gradient est équivalent à une pente de la surface virtuelle, pour une image virtuelle où x_p et y_p sont les coordonnées du pixel actuel où se trouve le joystick par rapport à la surface active de la F2T. La valeur de h est extraite via le canal vert du pixel. Lorsque la valeur de la hauteur donnée h est très élevée, cet effet permet de créer des murs virtuels infranchissables, simulant ainsi des obstacles physiques dans l'espace virtuel.

$$\vec{V}' = \vec{V} - \overrightarrow{\nabla} h(x_p, y_p) \tag{3.4}$$

C Effet de flux

Un autre effet intéressant mis en place par Gay *et al.* [169] est l'effet de flux. Quand le doigt arrive dans un champ de vecteur, il est entraîné dans la direction du vecteur avec l'amplitude du vecteur. Pour créer un effet de flux sur l'interface de la F2T, on crée manuellement un vecteur de déplacement en plusieurs points de la surface. La figure 3.5 présente un exemple de champ de vecteur. Les paramètres de contrôle, représentés dans la figure 3.5b permettent d'ajuster la norme du vecteur (p), de définir le rayon d'influence associé au point (D1), la distance seuil à partir de laquelle la norme décroît (D2) ainsi que la direction du flux par rapport à la courbe (angle).

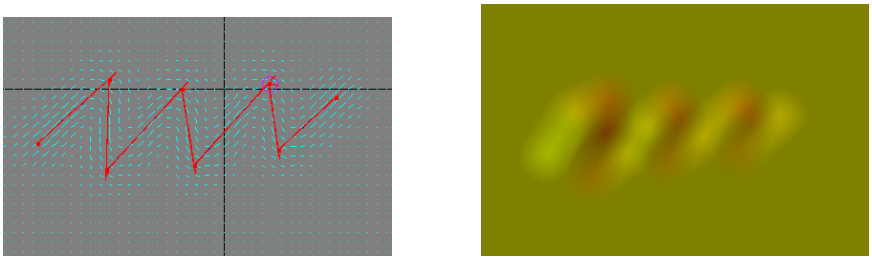
Ces paramètres sont ensuite utilisé pour créer une image virtuelle de flux (fig. 3.5a) où les composantes x et y de ces vecteurs sont encodées par le canal rouge et le canal vert respectivement (fig. 3.5c).

3.4 Modes d'interaction

Pour présenter une œuvre d'art à travers le dispositif haptique décrit précédemment, un mode d'exploration de type hiérarchique a été proposé [171]. Ce mode d'accès à l'information est décomposé en trois modes d'explorations différents : exploration guidée globale, exploration guidée détaillée et exploration sans guidage. Par ces modes d'exploration, on souhaite mettre en avant les règles de conception déduites des retours des participants dans l'étude exploratoire 1.

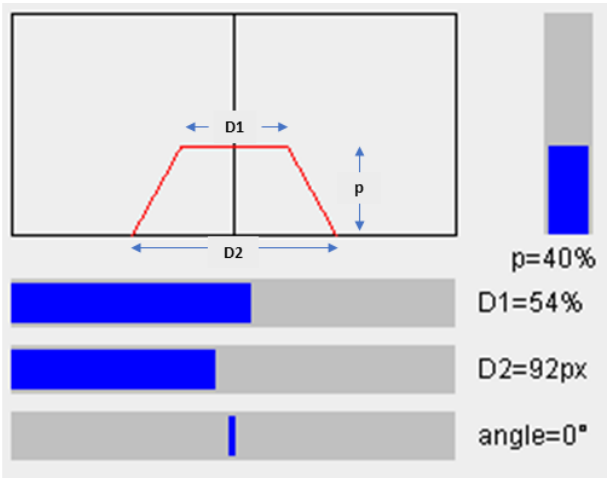
A Exploration guidée globale.

Cette étape correspond au premier niveau de compréhension de l'information, impliquant une exploration passive par les utilisateurs, surnommé le gist de niveau 1. Pendant cette phase, les utilisateurs perçoivent l'emplacement global des informations sémantiquement importantes. Ils posent leur index sur le joystick de la F2T pour suivre une trajectoire précédée par de courtes descriptions audio, avec des vitesses qui varient entre chaque point. La figure 3.7 présente un exemple de trajectoire annotée pour une image. Les vitesses attribuées vont de 1 à 500. Des courbes de bézier sont utilisées pour lisser la trajectoire entre deux points d'inflexion et une tangente, illustrés dans l'image par des flèches rouges.



(a) Champ vectoriel pour un effet de flux

(b) Image virtuelle générée pour l'effet de flux



(c) Interface de paramétrisation de l'effet de flux

FIGURE 3.5 – Présentation de l'interface pour créer un effet de flux



FIGURE 3.6 – Présentation hiérarchique de différents modes d'exploration d'une image virtuelle avec la F2T

B Exploration guidée détaillée.

Cette étape peut être décrite comme le deuxième niveau de compréhension de l'information, impliquant à la fois une exploration passive et active. Un utilisateur sélectionne un objet dans la scène en dirigeant le joystick vers l'élément qui l'intéresse puis appuie sur une touche du clavier. La figure 3.8 présente les sous-étapes inhérentes à cette méthode d'exploration.

Pour guider l'exploration, chaque objet est associé à un point d'attraction défini par un certain rayon d'influence. Dans la conception de dispositif haptique, les forces magnétiques sont indiquées comme un moyen efficace pour distinguer des zones interactives [80, 163]

À l'intérieur de ce rayon, une brève description verbale vient confirmer la découverte de l'objet. La force d'attraction appliquée modifie la vitesse initiale des moteurs, comme illustré dans l'équation 3.5. La nouvelle vitesse est alors calculée à partir de cette équation, où c représente la constante de force, et $\overrightarrow{PM}$ le vecteur reliant la position du joystick (P) au point d'attraction

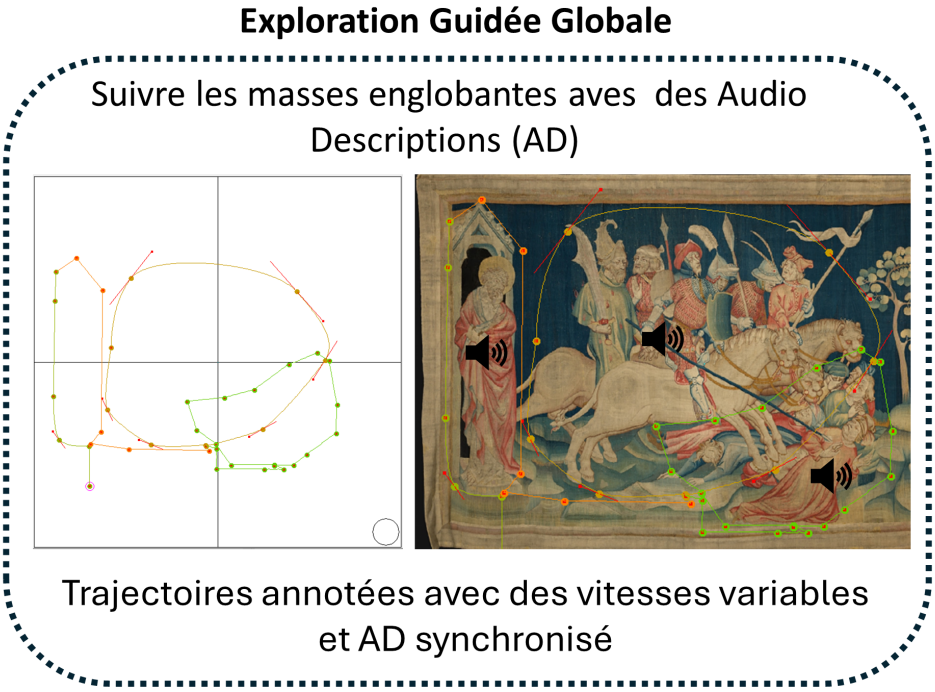


FIGURE 3.7 – Schéma illustratif du mode d'exploration guidée globale

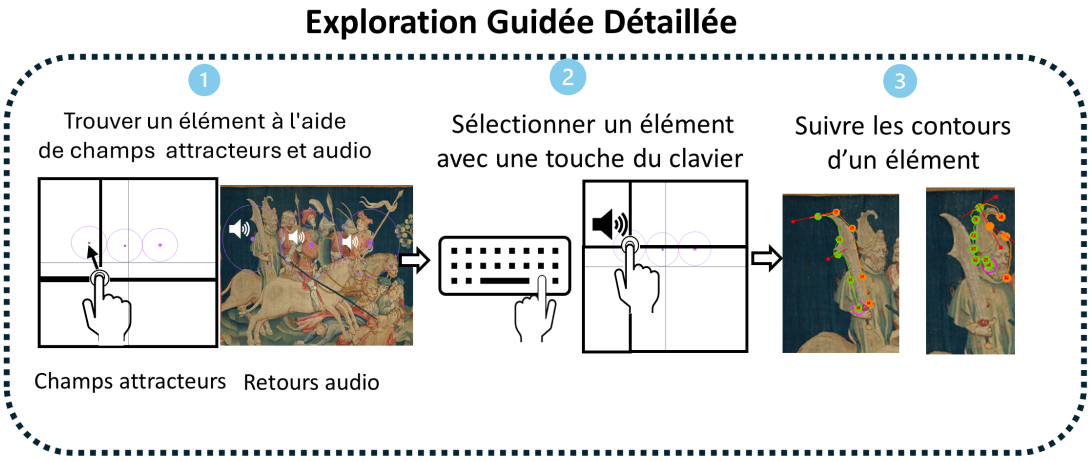


FIGURE 3.8 – Schéma illustratif du mode d'exploration guidée détaillée

(M).

$$\overrightarrow{V}' = \overrightarrow{V} + c\frac{\overrightarrow{PM}}{\|\overrightarrow{PM}\|} \tag{3.5}$$

Il s'agit d'un effet subtil conçu pour aider à sélectionner un objet dans une scène pour les non voyants. Une fois que l'utilisateur a sélectionné un objet, il revient à une perception passive des formes présentées. Ici, il est guidé pour suivre un chemin annoté, décrivant les contours des objets à travers de courtes descriptions audio.

C Exploration sans guidage.

Pour donner plus de liberté dans l'exploration, l'utilisateur est invité à choisir activement les éléments d'une scène, à explorer les contours de cet élément en utilisant à la fois le retour haptique et le retour audio. Les zones audio ont été définies en fonction de la densité des éléments présents dans la scène. Ces zones comportent des descriptions verbales brèves accompagnées d'effets sonores en boucle ou pas[173]. La figure 3.9 illustre un exemple d'images virtuelles pour les zones de retours audio et de retours haptique sur fond de l'image originale. Les retours haptiques correspondent à ceux présentés dans le paragraphe 3.3.

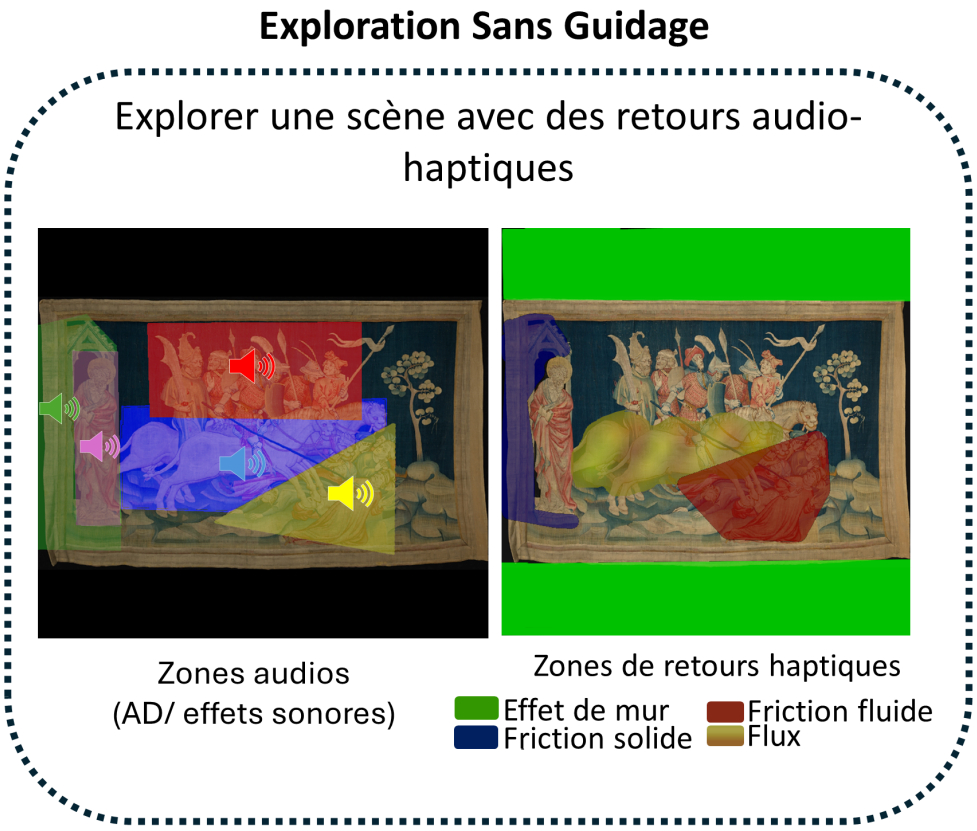


FIGURE 3.9 – Schéma illustratif du mode d'exploration sans guidage

4 Etude exploratoire 2 : Interaction avec un dispositif haptique

Cette section est une étude exploratoire qui est une étude qualitative. Elle est basée sur l'utilisation du dispositif haptique présenté ci-dessus, avec une application directe sur une œuvre d'art.

4.1 Introduction

Bien que des travaux ont exploré l'intégration multisensorielle, notamment à travers les modalités audio-haptiques, une attention insuffisante est accordée à l'accessibilité des œuvres d'art en 2D. Cette lacune est évidente dans les revues sur l'accessibilité pour les personnes aveugles et partiellement aveugles, où les solutions prédominantes reposent sur des répliques en 2.5D [174, 175].

À cet égard, la tablette à retour de force (F2T), un prototype haptique à 2 degrés de liberté, est utilisée. Dans des travaux précédents [169, 171], la conception de la F2T a été présentée et une méthodologie a été proposée ; cependant, aucune étude n'a été menée avec le dispositif en utilisant des images d'œuvres d'art avec le public ciblé. Ce travail vise à étudier comment le retour audio-haptique à travers une présentation d'interaction de type hiérarchique influence l'expérience des utilisateurs de la Tapisserie de l'Apocalypse. Similairement à la Tapisserie de Bayeux, la Tapisserie de l'Apocalypse est une œuvre conséquente composées de multiples scènes dans un style médiéval avec un nombre importants d'éléments représentés. Ainsi, il a semblé pertinent de tester les propositions d'exploration mis en place pour la F2T sur cette oeuvre. Cette étude vise à répondre à la sous-question de recherche suivante :

— Q2.4 : Comment l'utilisation du retour audio-haptique via une méthode d'exploration hiérarchique influence-t-elle l'expérience utilisateur de la Tapisserie de l'Apocalypse chez les participants aveugles, partiellement aveugles et non aveugles ?

Pour répondre à cette sous-question, ce travail est structuré comme suit. La Section 4.2 présente le profil des participants, en détaillant leur profil visuel et leur expérience avec les dispositifs audios et tactiles. La Section 4.3 décrit la méthodologie et le matériel utilisé, notamment les conditions d'expérimentations et les modalités audio-haptiques appliquées à la Tapisserie de l'Apocalypse. La Section 4.4 expose les résultats de l'étude et propose une discussion sur l'influence du retour audio-haptique et de l'exploration hiérarchique sur l'expérience des utilisateurs. Enfin, la Section 4.5 conclut l'étude en résumant les principaux enseignements et en ouvrant la discussion sur des recommandations pour l'accessibilité des œuvres d'art en 2D.

4.2 Profil des participants

Dix participants (3 PNA, 3 PA et 4 PPA ; âgés de 41 à 77 ans ; 4 femmes, 6 hommes) ont été recrutés grâce aux médiateurs du château d'Angers, notamment historienne en art et cheffe du service culturel du Domaine national du Château d'Angers, Catherine Leroi et son équipe, en accord

avec le responsable RGPD de l'Université de Rouen Normandie. Dans ce cadre les participants ont signé un formulaire de consentement éclairé et ont été informés par une lettre d'information de ce qui est attendu d'eux pour l'expérimentation. Ainsi une journée d'étude a été réalisée pour que les participants puissent tester les différents dispositifs devant la Tapisserie de l'Apocalypse. Dans la seconde étude pilote présentée ici, avec le dispositif haptique (F2T), c'est la Tapisserie de l'Apocalypse qui a fait l'objet de support pour les représentations créées dans le château d'Angers.

En effet, cette étude fait partie d'une expérimentation plus globale avec des tests sur deux autres dispositifs [141] : supports tactile en gaufrage et Puzzle 3D imprimé en résine.

Les gaufrages testés sont issus du livre Sensitinéraire sur la Tapisserie de l'Apocalypse [176]. Ce livre a été élaboré en 2015 collaboration avec le château d'Angers pour offrir des supports tactiles concernant quelques scènes de l'œuvre.

Le puzzle 3D correspond à une représentation d'une scène où différents éléments extractibles ont chacun été segmentés puis attribués une hauteur selon la méthode proposée par Redon *et al.* [108]. Contrairement à la F2T, ce sont des dispositifs non modulaires qui permettent une exploration bi-manuelle,

Afin d'éviter tout biais, nous avons sélectionné des scènes qui n'avaient pas été vues ni testées par les participants lors des tests avec les autres dispositifs. Ainsi, deux scènes ont été sélectionnées de la Tapisserie de l'Apocalypse (La Myriade des Cavaliers et La Grande Prostituée) pour cette nouvelle étude pilote, où chaque groupe a testé une image. Le groupe 1 est composé de deux PNA et trois PA. Le groupe 2, quant à lui, est composé d'un PNA et quatre PA.

TABLE 3.7 – Tableau présentant le profil visuel des participants ; PA : personne aveugle, PNA : personne non aveugle, PPA : personne partiellement aveugle ; F : genre féminin, M : genre masculin

Participants	Groupe 1	Participants	Groupe 2
P1	PA - cat V (F)	P6	PPA - cat IV (M)
P2	PPA - cat III (M)	P7	PPA - cat I (M)
P3	PNA (M)	P8	PA - cat V (M)
P4	PA - cat V (M)	P9	PNA (F)
P5	PNA (F)	P10	PPA - cat IV (F)

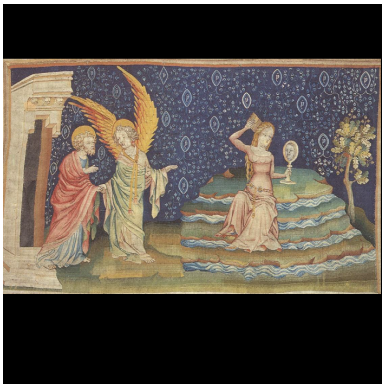
4.3 Méthode et matériel

La Tapisserie de l'Apocalypse est une œuvre d'art conséquente d'une longueur de 104m pour environ 4,50m de hauteur. Le récit de l'Apocalyse par St Jean est conté dans 74 scènes. Les scènes étudiées sont la scène S1 : "La myriade de cavaliers" (Fig.3.10a) et la scène S2 "La Grande Prostituée sur les eaux" (Fig.3.10b). Pour pouvoir être traitée, elles ont été redimensionnées selon l'espace de travail de la F2T, c'est-à-dire 20cm x 20cm et 701 pixels x 701 pixel.

Dans l'article [171], nous avons présenté une étude de cas décrivant deux modes principaux d'exploration intégrant le concept de gist tactile : un mode d'exploration guidée et un mode d'exploration sans guidage. Ces différents modes ont été présentés dans la section 3.4.



(a) Scène 26 - 2ème pièce - "La Myriade des Cavaliers"



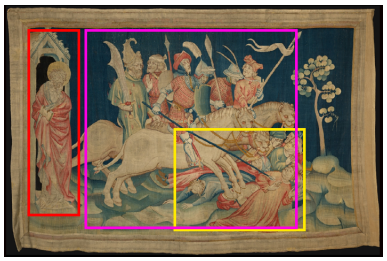
(b) Scène 64 - 5ème pièce -"La Grande Prostituée sur les eaux"

FIGURE 3.10 – Scène S1 (a) testée par le groupe 1 et Scène S2 (b) testée par le groupe 2.

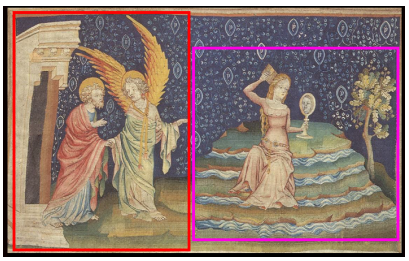
Cette étude élargit les travaux précédents en décomposant le mode d'exploration guidée, et en analysant l'expérience utilisateur sur le dispositif F2T en utilisant des scènes de la Tapisserie de l'Apocalypse. Elle reprend les modes de présentation d'exploration guidée globale, détaillée et sans guidage.

A Exploration guidée globale.

Cette méthode d'exploration introduit des chemins annotés et les descriptions verbales pour présenter de manière générale les éléments composant les scènes S1 et S2. La figure 3.11 présentent les boîtes englobantes désignant les éléments les plus sémantiquement importants pour chaque scène testée.



(a) "La Myriade des Cavaliers"



(b) "La Grande Prostituée sur les eaux"

FIGURE 3.11 – Boîtes englobantes des éléments les plus saillants des scènes S1 et S2

B Exploration guidée détaillée.

Dans le cadre de l'expérimentation proposée, un nombre limité d'éléments présentent une exploration guidée détaillée pour les scènes sélectionnées. Pour que les participants sélectionnent un élément en particulier, ils ont dû chercher ces éléments à l'aide d'attracteurs et de retours audios ces éléments. La disposition des attracteurs est présentée dans la figure 3.12. La scène S1 comporte 3 attracteurs rattachés à trois éléments sémantiquement important de la scène : le

Mercenaire d'Orient, Le Prince Noir, et le porteur d'Oriflamme (Fig. 3.12a). La scène S2, quant à elle, propose deux attracteurs situés sur Saint Jean et l'Ange, et la femme sur la colline (Fig. 3.12b). Pour chaque attracteur une description verbale permet d'affirmer le positionnement de l'utilisateur dans l'image.



(a) "La Myriade des Cavaliers" (b) "La Grande Prostituée sur les eaux"

FIGURE 3.12 – Positionnement des attracteurs pour les scènes S1 et S2

C Exploration sans guidage.

Dans ce mode d'exploration, les scènes sont présentées avec des retours haptiques et des retours audio. Le tableau 3.8 récapitule la modalité des retours selon les éléments de la scène affichés dans la figure 3.13. La figure 3.14 présente les effets de flux appliquées pour chaque scène. Chaque scène comportait quatre types d'effets haptiques : friction fluide, friction solide, effet de mur et effet de flux. Ces effets ont été appliqués à l'image pour renforcer sa valeur sémantique. Les effets de friction fluide et solide ont été modélisés en utilisant un coefficient de 50% pour le canal rouge et le canal bleu, correspondant à l'intensité moyenne sur la plage de 0 à 255. Les coefficients de friction ont été déterminés empiriquement pour minimiser les efforts nécessaires pour déplacer le joystick dans les zones données tout en conservant une force résistive légère (avec un coefficient de 100%, il est nécessaire d'exercer un effort important pour déplacer le joystick).

Les effets de mur sont désignés par le canal vert, fonction de hauteur, avec une valeur de coefficient maximale (100%) pour arrêter le mouvement, confinant ainsi l'exploration à un espace de travail plus réduit. Pour créer un effet de flux sur l'interface F2T, nous avons manuellement généré un vecteur de déplacement à plusieurs points dans l'espace de travail du F2T.

TABLE 3.8 – Tableaux comportant les effets présents dans les scènes S1 et S2. A : Audio, H : Haptique

Zones audio-haptiques	Modalités	Scène S1	Scène S2
AH_1	A	$Guérite + StJean$	$Guérite + StJean + Ange$
	H	*Friction solide*	*Friction solide*
AH_2	A	$Chevaux$	$Femme + Colline$
	H	*Flux*	*Flux+ Friction fluide*
AH_3	A	$Victimes$	-
	H	*Friction fluide*	-



(a) (b)

FIGURE 3.13 – Visualisation des zones audio-haptiques pour chaque scène étudiée



(a) (b)

FIGURE 3.14 – Visualisation des zones de flux pour chaque scène étudiée

4.4 Résultats & discussion

Les résultats obtenus suite à cette étude qualitative sont étudiés à travers un questionnaire dont les réponses sont sous la forme d'une échelle de Likert, des réponses à des questions libres posées. Nous avons aussi mesuré le temps consacré lors de l'exploration sans guidage pour pouvoir évaluer l'intérêt porté à l'exploration d'une part ou bien la difficulté de retrouver les éléments d'autre part.

A Retours des utilisateurs

La figure 3.15 présente les résultats au questionnaire présenté aux participants sous la forme de la médiane et des valeurs minimum et maximum. La liste des énoncés des questions est présentée dans le tableau 3.9.

Les participants ont été amenés à répondre à un questionnaire semi-structuré, comportant des énoncés à noter sur 5 (1 - "pas du tout" ; 2 - "un peu" ; 3 - "neutre" ; 4 - "assez" ; 5 - "beaucoup") et des questions libres pour chaque mode d'exploration testé.

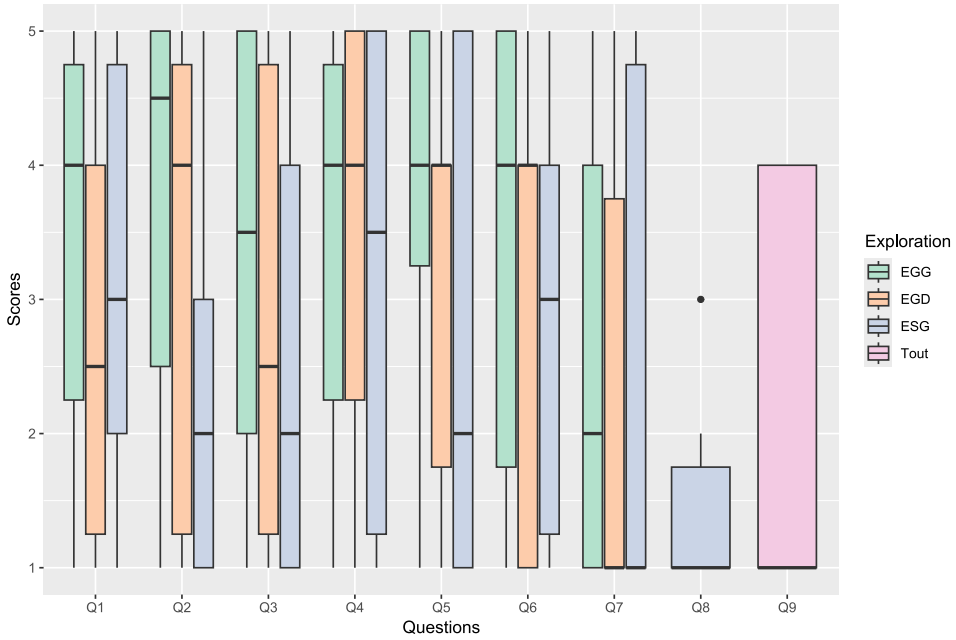


FIGURE 3.15 – Résultats échelle de Likert pour les questions du tableau 3.9 en lien avec l'Exploration Guidée Globale (EGG), l'Exploration Guidée Détaillée (EGD), l'Exploration Sans Guidage (ESG)

TABLE 3.9 – Enoncés liés au questionnaire de l'étude exploratoire 2.

Questions	Enoncés
Q1	J'ai apprécié la méthode d'exploration
Q2	La quantité d'information audio me suffit
Q3	La quantité d'information perçue tactilement me suffit
Q4	L'audio et le tactile s'intègrent bien ensemble
Q5	J'ai trouvé intéressant l'utilisation d'effets haptiques.
Q6	J'ai compris la scène présentée avec cette méthode d'exploration
Q7	Cette méthode d'exploration me suffit
Q8	J'ai ressenti différents effets haptiques
Q9	Les méthodes d'explorations sont répétitives

A.1 Appréciation des modes d'exploration

Le mode EGG ($med^{Q1}_{EGG} = 4.0[1;5]$) a reçu une appréciation plus élevée de la part des participants par rapport au mode EGD ($med^{Q1}_{EGD} = 2.5[1;5]$) et au mode ESG ($med^{Q1}_{ESG} = 3.0[1;5]$).

A.2 Quantité et forme de l'information audio

Les participants ont jugé suffisante la quantité d'information présentée pour l'exploration guidée globale ($med^{Q2}_{EGG} = 4.5[1;5]$) et l'exploration guidée détaillée ($med^{Q2}_{EGG} = 4.0[1;5]$). La forme de l'information audio diffère entre les modes : l'exploration guidée globale consiste en une présentation globale de la scène permettant de localiser les éléments entre eux, tandis que l'exploration guidée détaillée offre une description plus précise des éléments de la scène. En revanche, la quantité d'information audio pour l'exploration sans guidage a été jugée insuffisante ($med^{Q2}_{ESG} = 2.0[1;5]$).

Pour cette dernière, elle se limite à une description verbale simple et des effets sonores répétés en boucle. De plus, certaines zones, notamment le fond de la Tapisserie, n'avaient pas de retours audios, ce qui a été remarqué par les participants (P6 :*« Les zones silencieuses devraient être expliquées »*)

A.3 Perception tactile de l'information

La quantité d'information perçue tactilement a été jugée convenable pour l'exploration guidée globale ($med^{Q3}_{EGG} = 3.5[1;5]$), mais les résultats sont plus mitigés pour l'exploration guidée détaillée ($med^{Q3}_{EGD} = 2.5[1;5]$). Ces deux modes d'exploration reposent sur une exploration passive. Certains participants ont noté que le mode d'exploration guidée globale fournissait les dimensions des objets et leurs tailles relatives par rapport aux autres objets dans la scène. Certains participants PA ont été agréablement surpris de percevoir les contours en détail dans le mode de guidage détaillé. L'exploration guidée détaillée nécessitait une interaction via un attracteur qui n'était pas facilement perceptible. Ceci peut expliquer les notes plus basses pour ce mode d'exploration.

Les participants ont rencontré plus de difficultés à percevoir les retours haptiques lors de l'exploration sans guidage, jugeant la quantité d'information tactile peu suffisante ($med^{Q3}_{ESG} = 2.0[1;5]$). Les participants ont estimé que les effets de flux et d'attraction étaient plus prononcés par rapport aux effets de friction. D'autres participants ont remarqué que certains effets haptiques gênaient l'exploration (P1, P2 et P5). Le bruit des moteurs a perturbé certains participants, même avec des écouteurs. Certains effets ont été jugés difficiles à comprendre. En effet, les effets de flux n'étaient pas accompagnés d'explication audio.

A.4 Intégration audio-tactile

Toutefois, les participants ont estimé que les retours audio et tactiles s'intégraient bien ensemble pour les explorations guidées globale et détaillée ($med^{Q4}_{EGG/EGD} = 4.0[1;5]$). *Les étapes de guidage ont été travaillées pour que les descriptions audio annoncent les trajectoires fournies*. Les participants ont donné des avis positifs sur l'intégration des descriptions verbales et du retour haptique, les deux modalités améliorant l'exploration. Trois participants ont remarqué que les modes de guidage global et détaillé étaient complémentaires.

L'intégration audio-tactile pour l'exploration sans guidage a été jugée convenable ($med^{Q4}_{ESG} = 3.5[1;5]$). Trois participants ont mentionné le besoin de signaux sonores pour distinguer les objets (et leur description) - (P3) :*« Il n'y a pas de signal pour distinguer les éléments (dans le mode de guidage détaillé) »*. L'utilisation de repères sonores permettrait de mieux identifier les transitions entre les contours tracés.

A.5 Intérêt des effets haptiques

À la question Q5 sur l'intérêt des effets haptiques, l'exploration guidée globale et détaillée ont été bien reçues ($med^{Q5}_{EGG/EGD} = 4.0[1;5]$), en contraste avec l'exploration sans guidage

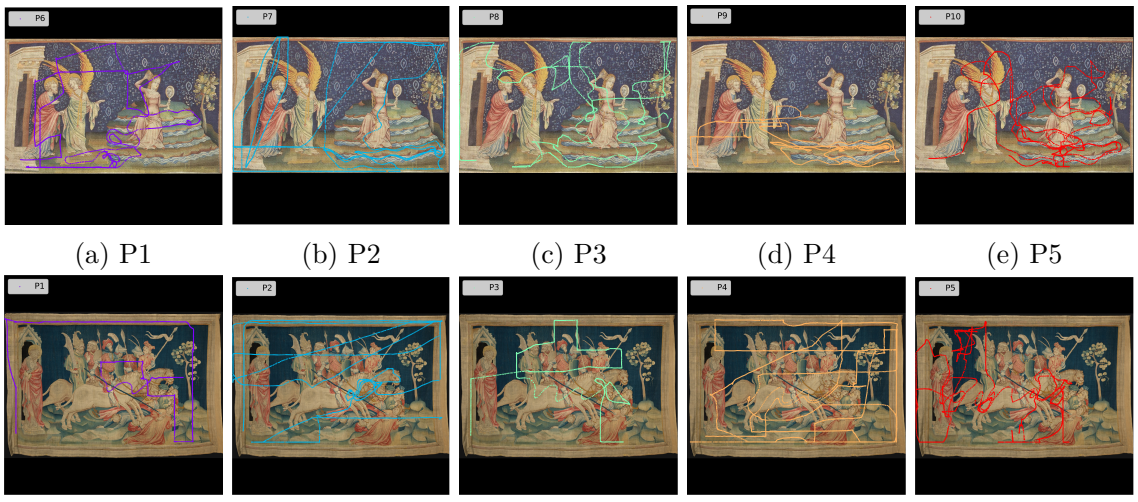
($med^{Q5}_{ESG} = 2.0[1;5]$), où les participants se sont sentis perdus. Il est possible de relier cela avec les retours des participants qui ont jugé ne pas avoir ressenti d'effets de friction ($med^{Q8}_{ESG} = 1.0[1;4]$). Effectivement, les effets de frictions étaient à 50%, contrairement aux effets de flux et de mur qui étaient réglés à 100%. Certains participants ont eu du mal à définir ce qu'est un retour haptique, alors qu'ils ont pu distinguer l'effet de flux et mur lors de l'exploration. Le mode d'exploration sans guidage nécessiterait une explication plus approfondie des effets haptiques, un participant P7 propose une phase d'apprentissage.

A.6 Compréhension des utilisateurs

En ce qui concerne la compréhension de la scène, les participants ont estimé qu'ils avaient bien compris la scène avec les modes d'exploration guidée globale et détaillée ($med^{Q6}_{EGG/EGD} = 4.0[1;5]$), tandis que les résultats sont plus neutres pour l'exploration sans guidage ($med^{Q6}_{ESG} = 3.0[1;5]$). Cependant, individuellement, ces méthodes d'exploration n'ont pas été jugées suffisantes ($med^{Q7}_{EGG} = 2.0[1;5]$ et $med^{Q7}_{EGD/ESG} = 1.0[1;5]$) pour l'expérience. Les participants ont souligné que les différents modes d'exploration se complètent, comme en témoigne leur avis sur la répétitivité des informations ($med^{Q9}_{Tout} = 1.0[1;5]$).

B Analyse temporelle du parcours des participants

Ce qui nous intéresse par la suite, est d'observer le comportement des participants pour l'exploration sans guidage. La figure 3.16 présente la trajectoire des participants pour chaque scène avec, en parallèle une estimation non paramétrique par noyau de la densité dans la figure 3.17.

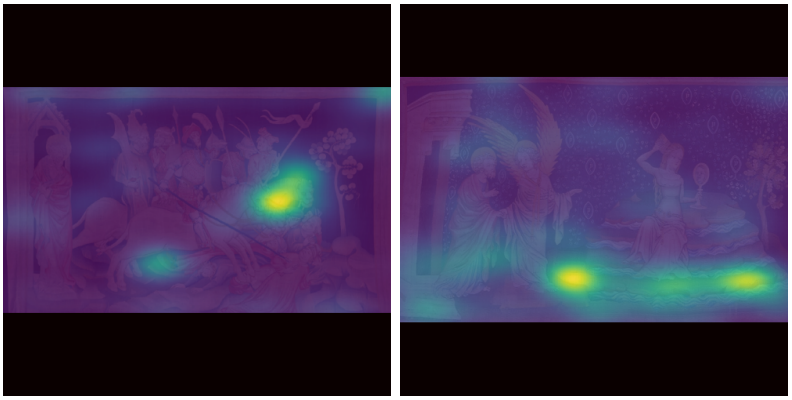


(a) P1 (b) P2 (c) P3 (d) P4 (e) P5

(f) P6 (g) P7 (h) P8 (i) P9 (j) P10

FIGURE 3.16 – Trajectoires des participants pour l'exploration de la scène S1 (haut) et la scène S2 (bas)

Ainsi, il est possible d'observer que les participants ont adopté des trajectoires à caractère principalement exploratoire. Cependant, en raison de la nature même de l'effet de flux, il apparaît



(a) S1 (b) S2

FIGURE 3.17 – Estimation de densité non paramétrique par noyau de la scène S1 et la scène S2

que leur exploration s'est concentrée majoritairement sur certaines zones spécifiques de la scène présentée.

Afin d'approfondir cette analyse, la figure 3.18 illustre le temps de présence normalisé (par le temps total) de chaque participant au sein des zones définies dans le tableau 3.8 et représentées dans la figure 3.13. Pour les deux scènes étudiées, il ressort que la zone AH_2 est celle où les participants ont passé le plus de temps, ce qui appuie leurs observations. En effet, l'effet de flux présent dans les zones AH_2 des scènes S1 et S2 a été perçu comme particulièrement « guidant ». Par conséquent, les autres effets haptiques ont été moins explorés et, de ce fait, moins perçus.

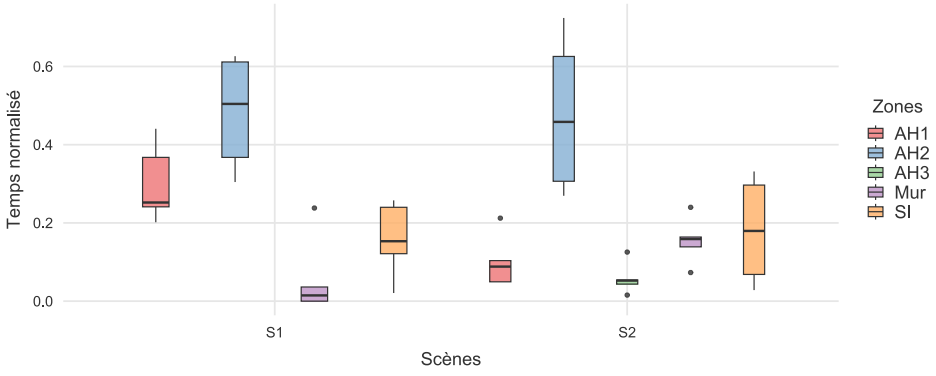


FIGURE 3.18 – Temps normalisé par zone audio-haptique explorée. *SI : Sans Information, Mur : contact aux murs simulés.

4.5 Conclusion

Cette étude a permis d'apporter des éléments de réponse à la sous-question de recherche : Q2.4 : Comment l'utilisation du retour audio-haptique via une méthode d'exploration hiérarchique influence-t-elle l'expérience utilisateur de la Tapisserie de l'Apocalypse chez les personnes aveugles, partiellement aveugles et non aveugles ?

Les participants ont généralement apprécié la manière dont le contenu de la tapisserie leur a été présenté à travers des modes d'exploration complémentaires (exploration guidée globale et détaillée). Certains ont été agréablement surpris par les détails fins des contours fournis par le mode de guidage détaillé. Cependant, étant donné qu'il s'agit d'une nouvelle technologie, les personnes ont besoin d'une phase d'apprentissage pour mieux comprendre les effets haptiques présents dans l'exploration sans guidage. En conclusion, les participants à cette étude exploratoire de la F2T ont valorisé les modes d'exploration guidés dans la présentation d'exploration hiérarchique. Les modes d'exploration étaient variés et complémentaires, facilitant l'exploration autonome des œuvres d'art en 2D par les visiteurs aveugles et partiellement aveugles des musées. Cependant, plusieurs limitations doivent être notées, notamment sur la taille de l'échantillon de participants pour chaque scène. L'un des points importants mis en avant sont les méthodes d'intégration de divers effets haptiques et audios pour l'exploration sans guidage.

5 Conclusion du chapitre

Ce chapitre vise à répondre à la question de recherche suivante : (Q2) comment présenter les œuvres d'art aux personnes aveugles et partiellement aveugles ? Avoir des retours audios et haptiques est indispensable dans la présentation d'une œuvre d'art à un public de visiteurs de musées. A la première étude, il est ressorti que les dispositifs actuels (audiodescription, thermogonflés, haut-reliefs) sont tous appréciés, et chaque dispositif possède un avantage pour présenter une œuvre d'art à un public inclusif.

Les retours des participants ont permis d'élaborer des recommandations pour le développement de nouvelles technologies d'accessibilité modulable. Sur cette base, un mode d'interaction hiérarchique a été intégré au dispositif haptique F2T. La seconde étude qualitative a montré que les méthodes d'exploration guidées, qui permettent aux utilisateurs de suivre les contours des œuvres de manière passive tout en recevant des descriptions verbales, sont particulièrement efficaces pour offrir une compréhension globale de l'œuvre, tout en permettant une exploration plus approfondie des détails spécifiques. Cela rejoint l'étude récente de Akar *et al.* [177] ayant montré qu'il est possible de reconnaître des formes à partir de mouvements passifs à travers un dispositif transcrivant des éléments visuels en retours audio couplé avec un système pantographe.

Par ailleurs, le mode d'exploration sans guidage, basé sur des retours audio-haptiques et leur exploration active, a été testé par des utilisateurs. Les résultats ont mis en évidence les limites de perception d'effets haptiques multiples pour une même scène et la nécessité d'une phase d'apprentissage. Dans ce cadre, il serait intéressant d'aller plus loin sur l'intégration de multiples retours haptiques pour la perception de scène pour faire en sorte que ces retours soient complémentaires et lisibles afin de mieux représenter la richesse de l'œuvre représentée. En conclusion, cette étude contribue à une meilleure compréhension des besoins des personnes aveugles et partiellement aveugles dans un contexte muséal, tout en identifiant les défis et opportunités associés à l'intégration des technologies multisensorielles dans la présentation des œuvres d'art.

4 Étude de la perception d'images de la Tapisserie de Bayeux

Sommaire

1 Introduction du chapitre 75
2 Base de données existantes 76
3 Base de données de la Tapisserie de Bayeux 78
3.1 Profil des participants 78
3.2 Conditions expérimentales 78
3.3 Résultats de la collecte de données 80
3.4 Discussion 95
4 Sélection de modèles 97
4.1 Introduction 97
4.2 Comparaison des méthodes de génération de cartes de saillance . . 97
4.3 Comparaison des méthodes de génération de parcours visuels . . . 102
5 Conclusion du chapitre 106

1 Introduction du chapitre

L'exploration d'une œuvre d'art par des personnes aveugles et partiellement aveugles suit généralement une exploration hiérarchique, allant d'une perception globale à une exploration détaillée [165]. Reproduire ce processus à travers des dispositifs haptiques et audio implique de sélectionner, structurer et transmettre les informations les plus pertinentes. Ce travail s'inscrit dans une démarche d'inclusivité, en facilitant l'accès aux œuvres d'art grâce à des représentations multisensorielles adaptées.

Dans ce contexte, l'objectif est de développer une approche automatisée permettant d'identifier et de hiérarchiser les éléments essentiels d'une scène visuelle. L'utilisation de l'intelligence artificielle (IA), appliquée à la génération de cartes de saillance et de parcours visuels, constitue ici une solution prometteuse pour guider l'adaptation des œuvres et réduire la dépendance à un travail manuel long et coûteux. L'étude présentée relève ainsi de la recherche appliquée, visant à exploiter les outils de vision par ordinateur pour résoudre un problème concret : rendre les images de la Tapisserie de Bayeux perceptibles par le dispositif haptique à retour de force F2T [169]. Ce dernier doit en effet pouvoir non seulement représenter des scènes, mais aussi simuler dynamiquement les interactions entre leurs éléments.

Trois questions de recherche structurent l'analyse :

— Comment identifier les éléments sémantiquement importants dans une œuvre d'art ?
— Quel est l'ordre d'importance dans lequel ces éléments sont explorés ?
— Dans quelle mesure les conditions de visionnage influencent-elles les parcours d'exploration ?

L'approche adoptée repose sur l'oculométrie appliquée à la Tapisserie de Bayeux, complétée par la comparaison de modèles d'IA pour la prédiction automatique de cartes de saillance et de trajectoires visuelles. L'objectif final est de sélectionner la méthode la plus adaptée pour généraliser ces prédictions à de nouvelles scènes, afin d'alimenter automatiquement le processus d'adaptation pour la F2T.

La méthodologie est organisée comme suit : (i) présentations de bases de données existantes pour la génération de cartes de saillance et de parcours visuels (section 2), (ii) constitution d'un jeu de données spécifique sur la Tapisserie de Bayeux (section et exploitation des données oculométriques collectées (section 3), et (iii) comparaison de modèles d'IA afin d'identifier ceux offrant la meilleure capacité de généralisation (section 4). La discussion et la conclusion soulignent les apports de cette démarche pour l'adaptation automatique d'œuvres au format haptique inclusif (section 5).

2 Base de données existantes

Dans la littérature, de nombreux jeux de données ont été développés pour l'évaluation des modèles de prédiction de la saillance visuelle sur les images, parmi lesquels les plus couramment utilisés figurent SALICON [178] et MIT1003 [179]. Pour l'analyse des parcours visuels, le jeu de données OSIE [180] constitue une référence majeure. Des bases de données plus spécialisées portant sur des œuvres picturales ont également été proposées, comme celle de Le Meur [181], axée sur les cartes de saillance, ou AVAtt [182], qui inclut à la fois des cartes de saillance et des parcours visuels.

Le tableau 4.1 présente des exemples de cartes de chaleur provenant des jeux de données SALICON, MIT1003 et de Le Meur. Le tableau 4.2 illustre des parcours visuels collectés auprès de quatre participants dans le cadre du jeu de données OSIE.

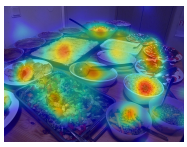
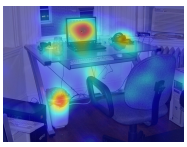
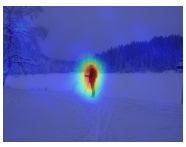
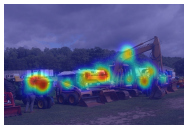
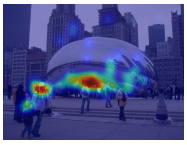
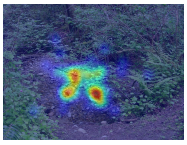
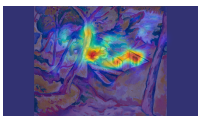
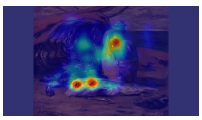
	Image 1	Image 2	Image 3	Image 4
Salicon[178]				
MIT1003 [179]				
Le Meur [181]				

FIGURE 4.1 – Tableau d'images illustrant les jeux de données Salicon [178], MIT1003 [179] et Le Meur [181].

Le jeu de données SALICON [178] contient 20 000 images de scènes naturelles, avec des cartes de saillance dérivées de mouvements de souris simulant l'attention visuelle humaine. MIT1003 [179] propose 1003 images observées pendant trois secondes en regard libre, avec des données recueillies via un eyetracker. OSIE [180] comprend 700 images de scènes naturelles, accompagnées de fixations oculaires et d'annotations sémantiques, également obtenues en condition d'exploration libre.

Ces jeux de données reposent principalement sur des scènes naturelles contemporaines, limitant ainsi leur pertinence pour l'étude d'images complexes issues du patrimoine. En effet, les représentations visuelles présentes dans les œuvres d'art peuvent se caractériser par une forte densité d'information, une composition narrative spécifique et une richesse symbolique. À ce jour, seules deux bases de données intègrent ce type d'images : celle de Le Meur *et al.* [181], composée de 151 peintures de différents courants artistiques (romantisme, impressionnisme, fauvisme, etc.), et AVAtt [182], qui élargit la couverture stylistique à 17 mouvements, y compris l'art abstrait, l'Ukiyo-e et le baroque. Toutefois, le jeu de données de Le Meur ne propose que des cartes de saillance, tandis qu'AVAtt, bien que plus complet, n'est pas encore disponible publiquement.

Dans cette étude, l'attention se porte sur un type d'image encore peu exploré dans ce domaine :



FIGURE 4.2 – Tableau d'images illustrants le jeu de données OSIE [180] pour 4 participants

les tapisseries et plus particulièrement la Tapisserie de Bayeux qui est une œuvre monumentale du XIe siècle relevant de l'art roman. D'une longueur de près de 70 mètres pour une hauteur de 50 centimètres, cette tapisserie se compose de 58 scènes à forte densité visuelle, mêlant narration historique et éléments décoratifs, dont la composition est similaire à une bande dessinée.

Le développement d'un jeu de données dédié à cette œuvre vise à combler les lacunes identifiées dans les corpus existants. De plus, contrairement aux bases disponibles, ce nouveau jeu de données s'appuie sur l'analyse de sous-parties localisées de l'œuvre, qui seront identifiées comme des images partielles d'une scène, afin d'étudier la perception visuelle à différentes échelles d'observation. Cette approche permet de tester l'influence de la granularité de la présentation – vue globale ou vue détaillée – sur les comportements visuels, notamment en termes de saillance perçue et de parcours oculaires.

La section suivante présente ce jeu de données intégrant à la fois des cartes de saillance et des parcours visuels, obtenus dans plusieurs conditions expérimentales contrôlées. L'objectif est d'évaluer si la perception de l'image varie significativement selon le niveau de zoom, mais aussi en présence de description verbale, et d'identifier les dynamiques attentionnelles associées à l'exploration d'une œuvre visuelle complexe et historiquement contextualisée.

3 Base de données de la Tapisserie de Bayeux

Cette section décrit le protocole suivi pour la constitution du jeu de données portant sur la Tapisserie de Bayeux. Elle présente dans un premier temps le profil des participants ayant pris part à l'étude (section 3.1), puis détaille la procédure expérimentale ainsi que le matériel utilisé pour la collecte des données (section 3.2). Enfin, un aperçu des images obtenues est fourni (section 3.3).

3.1 Profil des participants

Lors de ce test, 48 participants (33 hommes, 15 femmes), âgés de 21 à 64 ans, se sont portés volontaires. Étant donné que le système de calibration ne peut pas détecter correctement le regard des personnes partiellement aveugles, seuls des participants voyants ont pris part à l'expérience. Ils ont été invités à s'asseoir devant un ordinateur et à regarder des images de la Tapisserie de Bayeux. Cette expérience a été réalisée avec la validation du Règlement Général sur la Protection des Données (RGPD) de l'Université de Rouen Normandie. Avant le début de l'expérience, les participants ont été informés des objectifs de l'expérience. Ils ont été invités à remplir leur profil et ont signé un formulaire de consentement éclairé concernant l'utilisation des données collectées.

3.2 Conditions expérimentales

L'eyetracker utilisé est le Tobii Pro Fusion [1] a été utilisé pour cette expérience. Il fonctionne à une fréquence d'échantillonnage de 250 Hz. La distance moyenne entre l'appareil placé sous l'écran et les yeux des participants est d'environ 60 cm, avec une précision angulaire de $0.2°$.

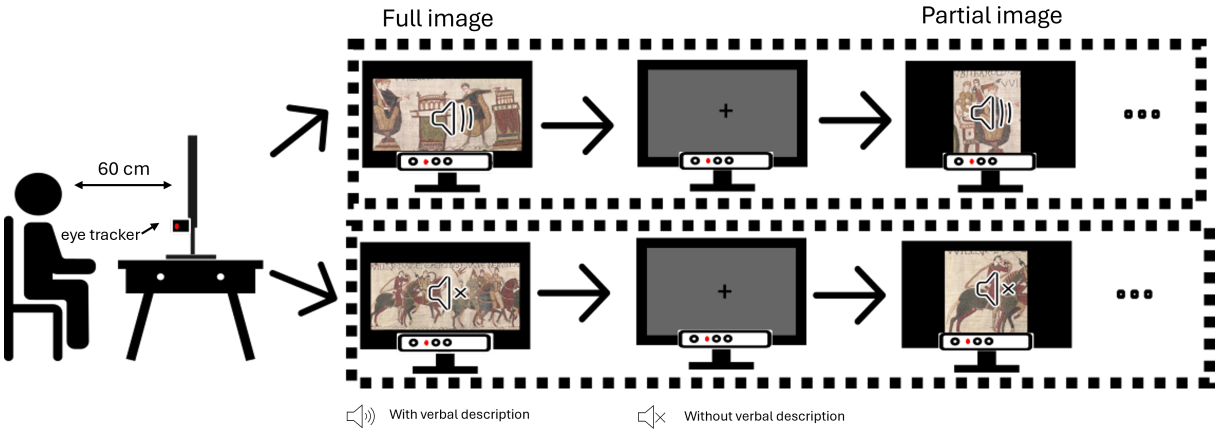


FIGURE 4.3 – Configuration expérimentale avec des exemples d'images affichées

La première phase de l'expérience consiste à calibrer l'Eye-tracker pour chaque participant. À cet effet, neuf points sont successivement affichés à différentes positions sur l'écran. Les participants sont invités à fixer leur regard sur les différents points. Une fois la calibration terminée, la deuxième phase commence avec la présentation des stimuli visuels. Des images de la Tapisserie de

1. Site de Tobii

Bayeux sont successivement affichées pendant 6 secondes chacune. Les participants regardent des images de diverses scènes, avec en moyenne 85 images par participant. Pour prévenir la fatigue oculaire, une pause de 2 minutes est proposée après environ 40 images.

La Figure 4.3 montre la configuration expérimentale des participants pendant la collecte des données. Entre chaque stimulus, un fond gris avec une croix au centre est affiché pendant 1 seconde. L'affichage commence par une image entière, qui est ensuite subdivisée en plusieurs images. La figure 4.4 illustre cette présentation hiérarchique. Chaque image est visualisée pendant 6 secondes, avec ou sans une description verbale. Un total de 314 images de 28 scènes de la Tapisserie de Bayeux ont été testées.

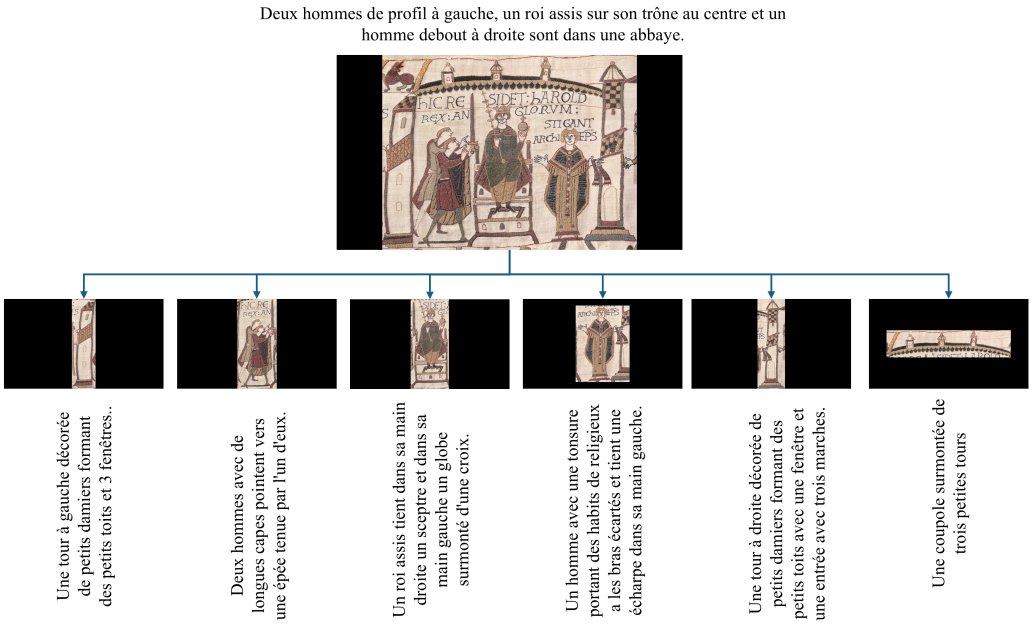


FIGURE 4.4 – Présentation hiérarchique de la scène 30 de la Tapisserie de Bayeux avec descriptions verbales.

Au cours de la collecte des données, quatre types de conditions ont été présentées :

— Observation d'images entières avec description verbale *(Ent_A)*.
— Observation d'images entières sans description verbale *(Ent_SA)*.
— Observation d'images partielles avec description verbale *(Part_A)*.
— Observation d'images partielles sans description verbale *(Part_SA)*.

Ces différentes conditions ont été présentées pour étudier comment le comportement visuel des participants est influencé par la présence ou l'absence de descriptions verbales et par les différentes manières dont les images ont été présentées.

À la fin de ces tests, les participants ont répondu à un questionnaire avec une échelle de Likert et quelques questions ouvertes.

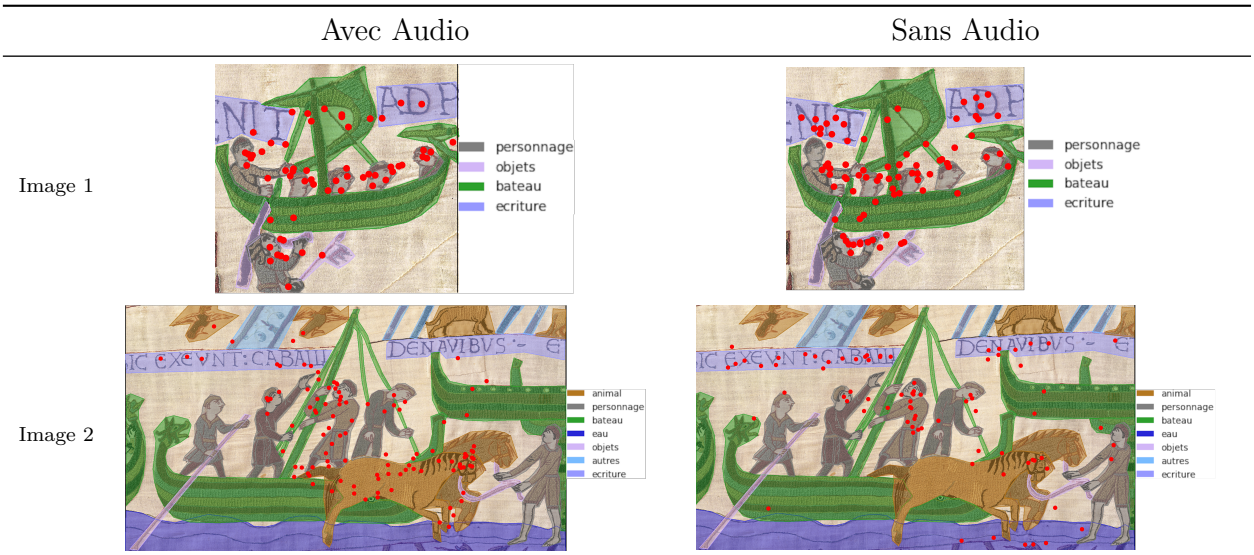


FIGURE 4.5 – Tableau d'images illustrant la dispersion des points de fixations dans les zones d'intérêts

3.3 Résultats de la collecte de données

Pour chaque image, des cartes de saillance moyennées (cartes de chaleurs) sur un ensemble de participants, et des parcours visuels pour chaque participant ont été obtenus.

A Zones saillantes

Cette section se charge de présenter les résultats liés aux zones de saillance par les observations qui ont été faites et une étude plus approfondie des zones d'intérêts.

Les cartes de saillance sont une technique de visualisation utilisée pour montrer la répartition de l'attention visuelle dans les données de mouvements oculaires. Elles mettent en évidence les zones sur lesquelles les participants ont concentré le plus leur attention, l'intensité de la carte de chaleur indique le niveau de concentration. Les points de fixation sont définis comme des périodes pendant lesquelles l'oœil reste immobile pendant plus d'un seuil donné (60 ms). Ces points indiquent que le système visuel humain se concentre sur des informations spécifiques dans la scène visuelle. La taille des zones observées dans les cartes de saillance correspond au temps passé à se concentrer sur ces régions. Les zones avec plus de points de fixation sont plus grandes, ce qui indique une attention visuelle plus élevée.

La Figure 4.5 présente un exemple d'image annotée avec des régions d'intérêts. Une observation avec description verbale à gauche et une observation sans description verbale à droite sont données pour un ensemble de participants. Les points de fixation sont représentés par des points rouges.

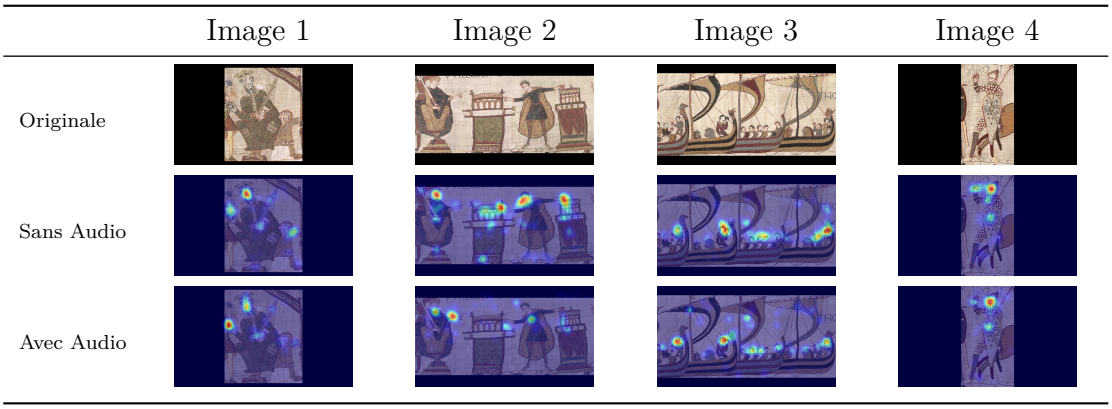


FIGURE 4.6 – Tableau de comparaison des cartes de saillance pour des participants ayant observé plusieurs scènes de la Tapisserie de Bayeux dans des conditions sans audio et avec audio. Les cartes sont moyennées selon le nombre de participants ayant observé les images dans les mêmes conditions.



FIGURE 4.7 – Tableau de comparaison des cartes de saillance pour des participants ayant observé une scène de la Tapisserie de Bayeux dans des conditions sans audio et avec audio. Les cartes sont moyennées selon le nombre de participants ayant observé cette image dans les mêmes conditions.

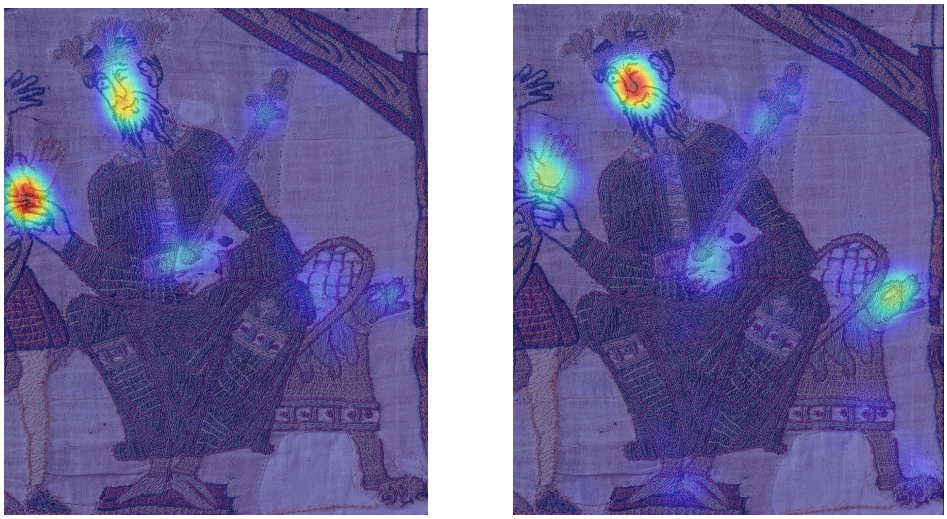
A partir de ces points de fixations, il est possible de générer des cartes de saillance. La figure 4.6 présente un exemple de cartes de chaleurs qu'il est possible de récupérer pour des scènes issues de la Tapisserie de Bayeux.

Les images montrées mettent notamment en avant que les participants de l'étude ont tendance à observer principalement les personnages, comme il est possible d'observer dans la figure 4.7.

Cette observation est appuyée par l'étude des zones d'intérêt, notamment en calculant la somme cumulée des points de fixations dans chaque zone d'intérêts pour l'ensemble des images testées. Afin d'identifier les éléments les plus fréquemment observés, le nombre de points de fixation dans les régions d'intérêt a été extrait à partir de 135 images, chacune ayant été présentée dans des conditions avec et sans audio, testées par les participants.

Ces zones d'intérêts ont été définies manuellement. Chaque zone est annotée avec un nom de catégorie suivant : personnage, monument, animaux, écriture, objets, nature, bateau, arme et autres. La catégorie "autres" désigne toutes les formes ornementales situées dans les bandes supérieures et inférieures encadrant la bande principale de la Tapisserie de Bayeux.

La figure 4.9 montre la somme cumulée pour chacune de ces catégories pour les conditions avec ou sans audios. Cette figure appuie notamment l'observation faite dans la figure 4.6 où les zones les plus observées appartiennent à celles annotées comme "personnages".



(a) Avec audio (b) Sans Audio

FIGURE 4.8 – Observation en détails des zones saillantes pour la scène 1 de la Tapisserie de Bayeux

Ceci rejoint la littérature, puisque les êtres humains depuis l'enfance, tendent à porter leur attention sur les visages principalement [183-186]. Autres que cette catégorie les participants se sont concentrés sur les monuments, les animaux et les objets, et ceci qu'importe les conditions sonores.

Comme observé dans la figure 4.8, les sous-zones des personnages principalement observées correspondent aux visages et aux mains.

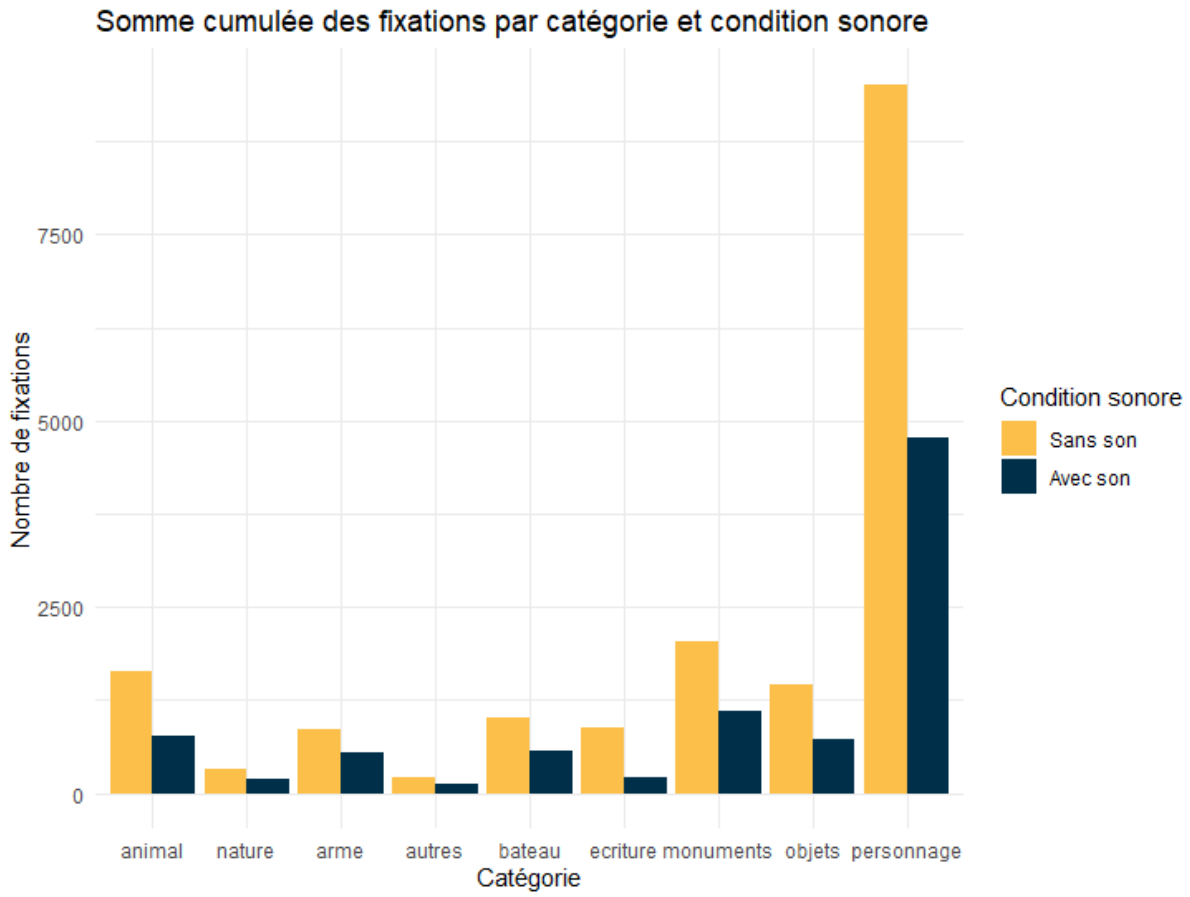


FIGURE 4.9 – Somme cumulée des points de fixations pour chaque catégorie pour les conditions sonores : Avec et Sans audio

B Parcours visuel

Les parcours visuels illustrent des mouvements oculaires sur l'écran, par les points de fixations et les saccades, suivant l'ordre logique de traitement de l'information par le participant. Ces parcours visuels offrent un aperçu du chemin de l'exploration visuelle à mesure que les participants observent un stimuli.

Les saccades font référence aux mouvements rapides des yeux qui se produisent lorsque l'œil passe d'un point de fixation à un autre. Ces mouvements sont souvent observés lorsqu'un participant change de point de focus ou cherche de nouvelles sources d'information dans une scène. Les mouvements saccadiques peuvent donner un aperçu de la façon dont les participants naviguent dans une scène visuelle pour recueillir des informations.

B.1 Étude de la direction

Pour l'étude de la direction 135 images de 11 scènes ont été retenues. Ces images ont été retenues car elles correspondent à celles observées dans les deux conditions expérimentales, avec et sans description verbale. La figure 4.10 illustre des parcours visuels des participants de l'étude pour une image issue de la scène 25 de la Tapisserie de Bayeux. Cette scène représente Harold prêtant serment à Guillaume le Conquérant. Ainsi, les participants A et B (P_A et P_B dans la Fig. 4.10) ont observé des images de cette scène sans description verbales, tandis que les participants C et D (P_C et P_D dans la Fig. 4.10) ont observé les mêmes images avec une description verbale. Il est possible de remarquer ici la variabilité des parcours visuels selon les participants. Le balayage reste principalement horizontal pour les deux groupes d'observations.

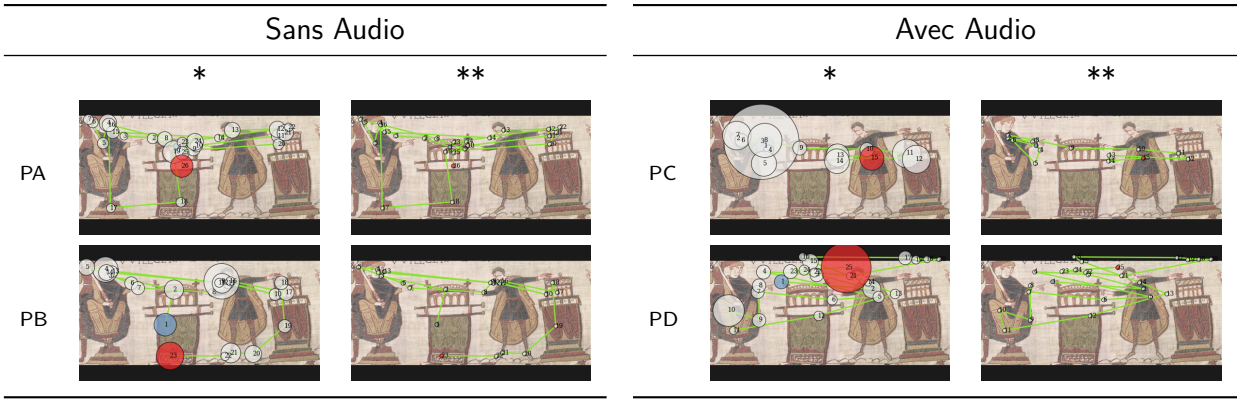


FIGURE 4.10 – Parcours visuels lors de l'observation d'une scène de la Tapisserie de Bayeux, avec et sans audio. Les cercles bleus indiquent le premier point de fixation, les rouges le dernier, et les lignes vertes les saccades. * : taille des cercles proportionnelle à la durée des fixations ; ** : sans indication de durée.

Une analyse a été réalisée afin d'évaluer si les directions du regard étaient affectées par les conditions de visionnage : format de l'image et présence de descriptions verbales. Les dimensions de l'affichage utilisées durant l'étude étaient de 1920 $\times$ 1080 pixels. Pour les images n'occupant pas l'intégralité de l'écran en raison de rapports d'aspect différents, des bordures noires ont été ajoutées. Les ratios correspondant à l'espace réellement occupé par les images en largeur et en hauteur ont ainsi été calculés.

En utilisant l'algorithme des K-moyennes (k-means clustering) appliqué à ces ratios (largeur W et hauteur H), il a été possible de regrouper l'ensemble des images en trois clusters distincts.

La figure 4.11 illustre la distribution des images au sein des différents clusters. Le tableau 4.1 présente les valeurs minimales et maximales des ratios calculés en largeur et en hauteur. Par ailleurs, des exemples représentatifs des images réparties dans les clusters 0, 1 et 2 sont fournis dans la figure 4.12.

Le cluster 0 contient principalement des images de petit format, tandis que le cluster 2 regroupe les images présentant les dimensions les plus larges et les plus hautes. Le cluster 1 correspond à des formats intermédiaires, caractérisés par des images plus hautes que larges.

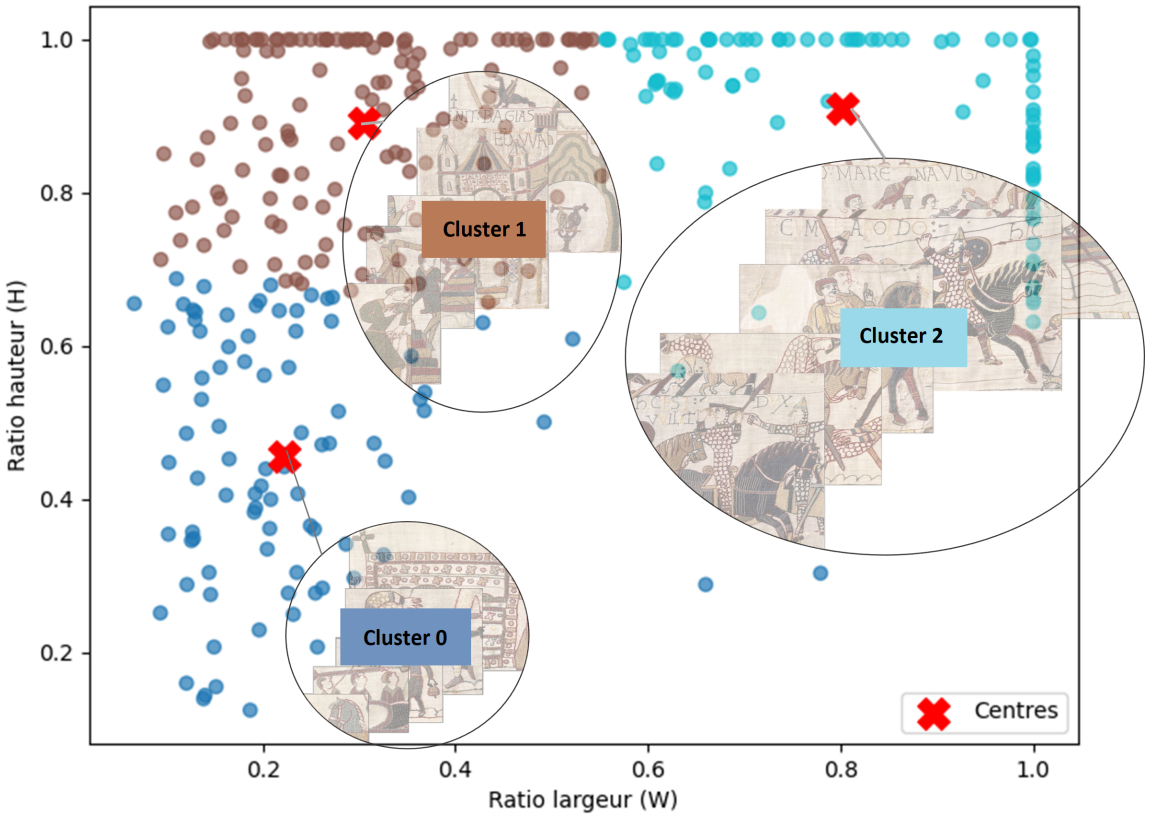


FIGURE 4.11 – Partitionnement des images selon la méthode des K-moyennes pour trois clusters définis à partir du ratio sur la largeur (W) et la hauteur (H).

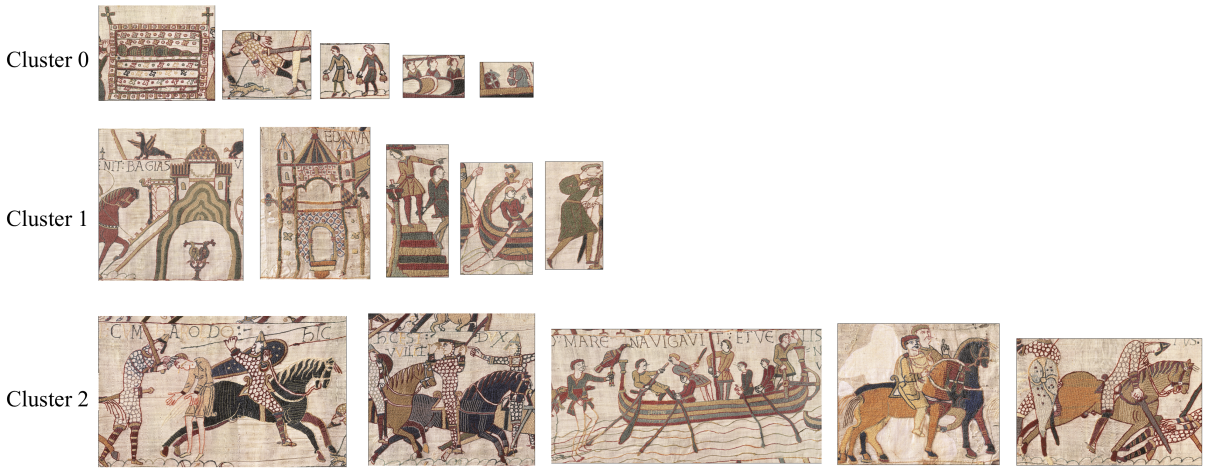


FIGURE 4.12 – Exemples d'images appartenant aux cluster 0, 1 et 2.

	ratio W		ratio H	
Clusters	min	max	min	max
0	0.07	0.78	0.13	0.69
1	0.09	0.55	0.66	1.00
2	0.56	1.00	0.57	1.00

TABLE 4.1 – Tableau des valeurs minimales et maximales des ratios calculés en largeur (W) et en hauteur (H) pour les images affichées dans une résolution de 1920×1080 pixels

La figure 4.13 présente les histogrammes des angles calculés à partir des parcours visuels des participants ayant observé les images avec ou sans description verbale. La figure 4.14 montre les histogrammes des angles calculés en fonction du format des images (Entières ou partielles), sur la base des clusters identifiés. Les histogrammes circulaires (Fig. 4.13 et Fig. 4.14) représentent les fréquences relatives plutôt que les fréquences absolues, afin de permettre la comparaison de la forme des distributions indépendamment du nombre total d'observations. En effet, le cluster 1 comprend 72 images, contre 30 pour le cluster 0 et 33 pour le cluster 2. Par ailleurs, les images issues du cluster 0 sont de plus petite taille.

D'après la figure 4.13, les participants ont plus tendance à faire des parcours horizontaux quels que soient les retours audios.

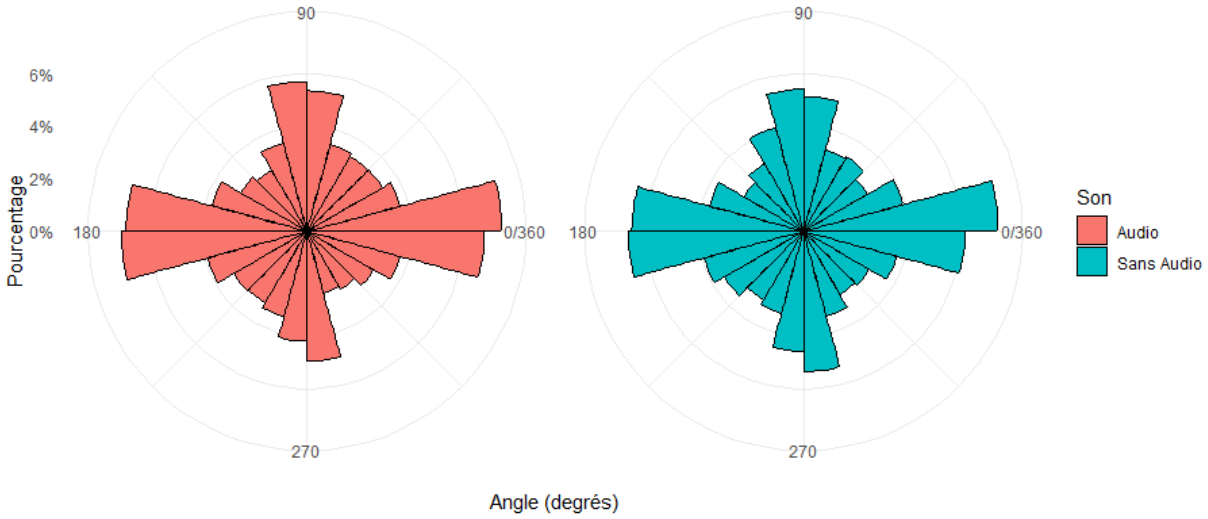


FIGURE 4.13 – Histogrammes circulaires des angles calculés en degrés pour les conditions avec ou sans retour audio, exprimés en fréquences relatives.

Les histogrammes de la figure 4.14 montrent que les angles de saccades des parcours sont majoritairement horizontaux pour les images de petit format (cluster 0) et de grand format (cluster 2). Le cluster 1, quant à lui, présente des angles plus prononcés sur l'axe vertical que l'axe horizontal.

Le test de Watson U^2 [187, 188] est utilisé pour vérifier si les angles sont uniformément répartis ou s'ils présentent une distribution non uniforme.

Ainsi, pour les images présentées avec ou sans retours audios, les distributions angulaires ont

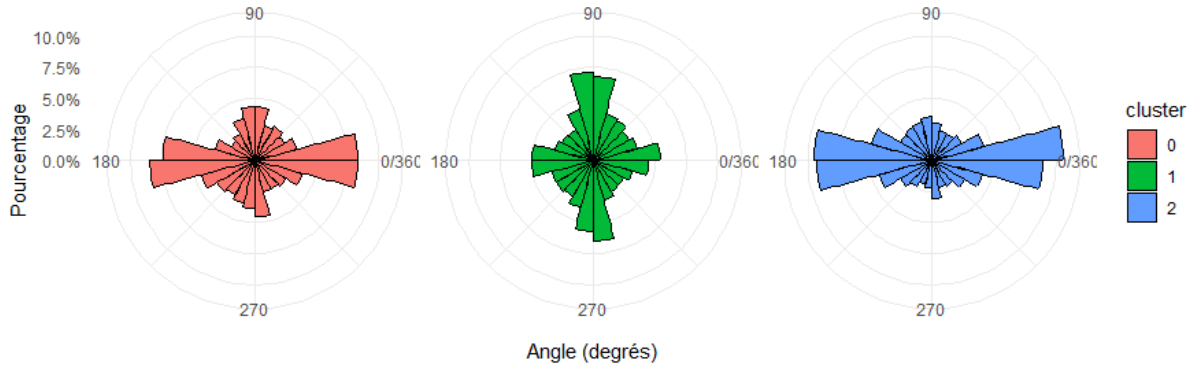


FIGURE 4.14 – Histogrammes circulaires des directions des parcours visuels selon la condition de format de l'image, exprimés en fréquences relatives.

été analysées afin de détecter toute déviation par rapport à une distribution uniforme sur le cercle, sans supposer de direction privilégiée.

La statistique U^2 renseigne sur la concentration des angles autour d'une direction. Une valeur élevée indique une concentration forte, tandis qu'une valeur faible reflète une concentration faible. La valeur p permet d'évaluer la significativité du test : une valeur inférieure à 0.05 indique une déviation significative par rapport à l'uniformité, tandis qu'une valeur supérieure suggère l'absence de différence significative pour les groupes étudiés.

Les résultats montrent que, pour la condition avec retour audio, le test Watson fournit une statistique $U^2 = 70.07$ avec une $p_{value} < 0.01$, ce qui met en évidence une déviation significative par rapport à une répartition uniforme. Les angles ne sont pas uniformément répartis et présentent une direction préférentielle marquée pour la condition avec retour audio. Pour la condition sans retour audio, la statistique est modérée ($U^2 = 44.95$, $p_{value} < 0,01$), et révèle une non-uniformité. Les angles présentent ainsi une concentration notable, bien que moins prononcée que dans la condition avec retour audio.

En s'intéressant au format de l'image, c'est-a-dire, s'il s'agit d'une image appartenant au cluster 0 (petite), 1 (intermédiaire) ou 2 (grande), on observe la même tendance d'angles non uniformément répartis avec une statistique plus élevée pour les images appartenant au cluster 2 ($U^2 = 191.49$, $p_{value} < 0,01$) que pour les images du cluster 1 ($U^2 = 41.55$, $p_{value} < 0,01$) et du cluster 0($U^2 = 39.31$, $p_{value} < 0,01$).

Comme l'illustrent les histogrammes circulaires (Fig. 4.14), on observe que les directions horizontales sont davantage privilégiées que les directions verticales pour les images des clusters 0 et 2, tandis que les directions verticales semblent plus marquées pour le cluster 1. Enfin, que l'observation des images soit accompagnée ou non d'une description verbale (Fig. 4.13), les directions horizontales demeurent prédominantes, suivies des directions verticales.

B.2 Influence des conditions sur la similarité des parcours

Dans la section précédente, l'analyse portait sur un nombre restreint d'images pour l'ensemble des participants, sans évaluer de manière combinée les différentes conditions ni tenir compte des variations entre participants. Étudier la similarité des parcours visuels permettrait ainsi d'examiner l'influence conjointe des conditions de visionnage à l'aide de mesures de similarité.

Afin de mesurer la dissimilarité des données des parcours de regard des participants, la méthode DTW (Dynamic time warping) [189] a été utilisée. C'est une mesure permettant la comparaison de séries temporelles.

Soient deux séries temporelles S et T, de longueur respective n et m, correspondant aux séquences de points de parcours visuels. Pour calculer la mesure DTW, les deux séries temporelles constituent une matrice de distances nommées $D_{(S,T)}$.

La matrice de distances euclidienne $D_{(S,T)}$ est définie par :

$$D_{(S,T)}[i,j] = |s_i - t_j| \tag{4.1}$$

avec $0 \leq i \leq n$ et $0 \leq j \leq m$.

En calculant la matrice des distances cumulées $DTW_{(S,T)}$, le chemin de déformation optimal reliant les deux séries dans une matrice de distance est recherché, de manière à ce que la somme des distances locales entre points appariés soit minimale. Ainsi la mesure DTW est définie par :

$$DTW(i,j) = \begin{cases} DTW(0,0), & \text{si } i = j = 0 \\ D(i,j) + min \begin{cases} DTW(i-1,j), \\ DTW(i,j-1), \\ DTW(i-1,j-1) \end{cases} & \text{sinon} \end{cases} \tag{4.2}$$

MultiMatch [190] est une métrique dépendant de zones d'intérêt (AOI - Area of Interest) pour évaluer la similarité des trajectoires de balayage selon cinq dimensions : la durée (MM-Dur), la direction (MM-Dir), la longueur (MM-Len), la forme (MM-Shape) et la position (MM-Pos). La métrique Multimatch a été calculée pour évaluer l'effet des conditions sur la similarité lors de l'observation d'images par les différents participants.

— MM-Dur mesure la similarité des durées de fixation entre les trajectoires de balayage prédit et du balayage réel.
— MM-Dir évalue la différence angulaire entre les vecteurs de saccades dans les trajectoires de référence et les trajectoires prédites.
— MM-Len examine dans quelle mesure les longueurs des vecteurs de saccades correspondent entre les deux trajectoires prédite et réelle.
— MM-Shape capte la similarité structurelle globale des trajectoires en analysant les différences entre les vecteurs de saccades.

— MM-Pos quantifie la précision des positions de fixation en comparant les localisations du regard prédites et observées.

Les scores des sous-métriques de MultiMatch varient de 0 à 1, les valeurs plus élevées indiquant une meilleure correspondance entre les trajectoires de balayage prédites et réelles pour la dimension donnée.

Dans le cadre de l'analyse de la similarité des parcours de fixations des participants, les effets des conditions comme le format d'image, selon leur dimension, et la présence de descriptions (Avec/Sans audio) sont évalués.

Pour la présentation des résultats, une nomenclature est utilisée afin de désigner les moyennes des différentes métriques (DTW, Multimatch). M correspond à la moyenne de la mesure considérée, l'exposant (e.g., $^{MM_{dur}}$ ou ^{dtw}) indique la métrique désignée. L'indice inférieur précise la condition expérimentale : SA pour sans audio et A pour avec audio, tandis que le chiffre (0, 1 ou 2) renvoie au cluster d'images correspondant. Ainsi, M_{A2}^{dtw} correspond à la moyenne de la mesure DTW pour le cluster dans la condition sans audio.

Une analyse statistique a été réalisée à l'aide de R (R Core Team, 2023) [2] et du package BRMS [191] (Bayesian Regression Models using 'Stan') pour examiner la mesure DTW et la métrique Multimatch sur 135 images réparties en trois groupes, selon le ratio entre les dimensions réelles de l'image et celles de l'image affichée, plus petite que le cadre de l'écran.

Le package BRMS permet d'ajuster des modèles bayésiens généralisés (linéaires et non linéaires), multivariés et multiniveaux.

Pour la présentation des résultats, b désigne le coefficient moyen des effets fixes estimés pour chaque métrique (DTW, Multimatch), l'exposant indiquant la métrique (e.g., $^{MM_{dur}}$ ou ^{dtw}).

Dans le cas de l'étude, le calcul de DTW et des sous-métriques Multimatch concernent les trajectoires d'un $Participant_1$ et d'un $Participant_2$.

B.2.1 DTW Dans un premier temps, la moyenne et l'écart-type ont été calculés pour les valeurs de DTW. Pour vérifier si les interactions entre les différentes conditions influencent les valeurs calculées de DTW. On propose d'analyser ces résultats à l'aide d'une modélisation hiérarchiques à effets mixtes.

La mesure DTW a été modélisée en utilisant la distribution gamma avec un lien logarithmique. Le modèle suivant a été utilisé :

$$DTW = condition_{son} * cluster_{format} + (1|Participant_1) + (1|Participant_2) + (1|Image) \quad (4.3)$$

Dans ce modèle, DTW représente la variable dépendante, tandis que $condition_{son}$ et $cluster_{format}$ sont considérés comme des effets fixes. Les termes $(1|Participant_1)$,$(1|Participant_2)$ et $(1|Image)$ correspondent aux effets aléatoires sur l'ordonnée à l'origine ("intercepts aléatoires"), permettant

2. R Core Team : langage R

de modéliser la variabilité associée à chaque participant et à chaque image.

Dans l'approche bayésienne, la significativité des effets est évaluée à l'aide d'intervalles de crédibilité (IC). Un intervalle de crédibilité correspond à la plage de valeurs dans laquelle un paramètre du modèle (par exemple un coefficient d'effet fixe ou une variance d'effet aléatoire) a une probabilité donnée (ici 95%), de se trouver, au regard de sa distribution a posteriori. Cette distribution a posteriori représente la probabilité du paramètre compte tenu des données observées et des distributions a priori spécifiées dans le modèle bayésien. Lorsqu'il inclut zéro cela indique l'absence d'un effet significatif. Ici, l'ordonnée à l'origine (appelée "intercept") représente les valeurs associées à la condition de référence, à savoir les images appartenant au cluster 0 avec la condition présentant de l'audio. Par conséquent, toutes les comparaisons sont faites par rapport à cette condition de base. Les résultats obtenus pour cette analyse sont présentés dans le tableau 4.2.

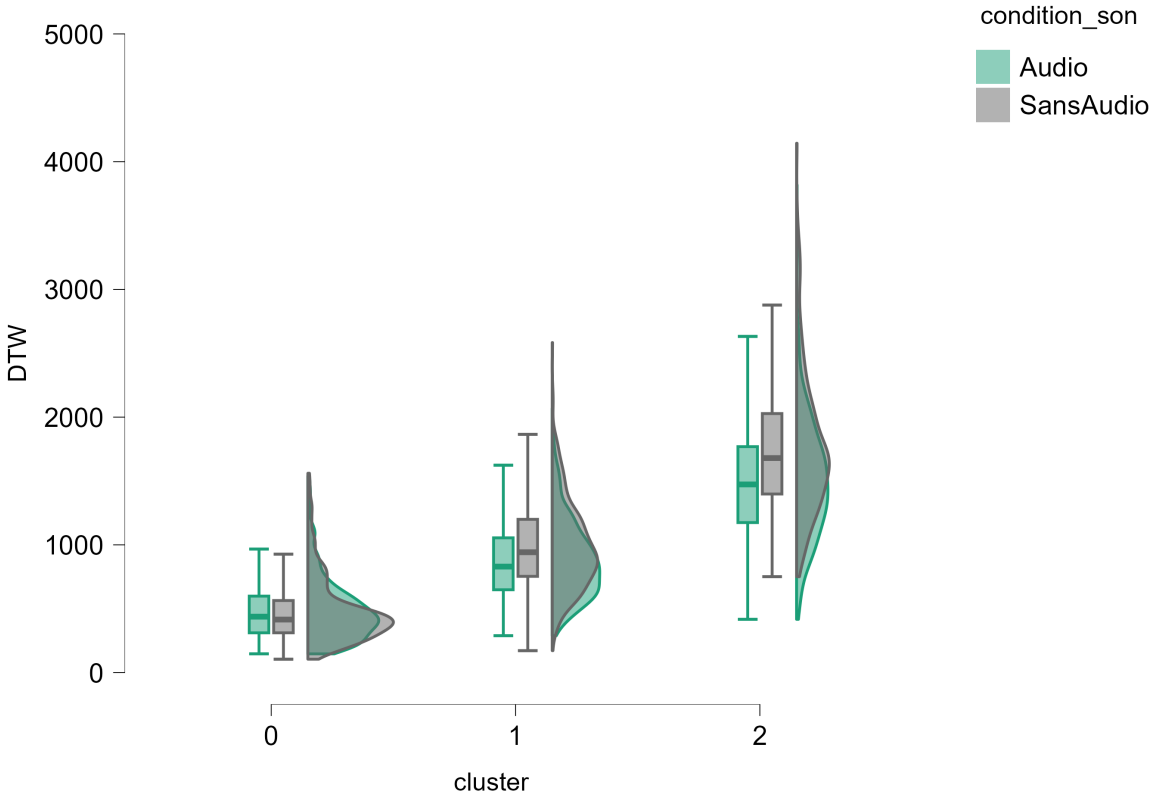


FIGURE 4.15 – Boîtes à moustache des résultats pour la mesure DTW

La figure 4.15 présente les résultats DTW calculées selon les conditions de son (avec et sans description verbale), et les conditions sur les images par rapport à leur format (appartenance au cluster 0, 1, 2). Les mesures DTW, calculées pour les différentes conditions audio et de format de l'image, montrent dans un premier temps que le cluster 2 ($M^{dtw}_{SA2} = 1788.32 \pm 578.85$ vs $M^{dtw}_{A2} = 1501.32 \pm 479.87$) et 1 ($M^{dtw}_{SA1} = 994.25 \pm 350.66$ vs $M^{dtw}_{A1} = 878.94 \pm 304.81$) présentent le plus de dissimilarités, notamment pour la condition sans audio comparé à la condition avec audio. D'après l'analyse bayésienne, une augmentation a lieu lors du changement du cluster 0 vers le cluster 1 ($b^{dtw} = 0.60\ [0.47 - 0.73] 95\% IC\%$) et le cluster 2 ($b^{dtw} = 1.19\ [1.04 - 1.33] 95\% IC$). Tandis que les différences sont moins notables pour les conditions avec ou sans audio pour le cluster 0 ($M^{btw}_{SA0} = 487.14 \pm 266.55$ vs $M^{dtw}_{A1} = 490.88 \pm 255.81$). En effet, la présence ou l'absence d'audio pour le cluster 0, n'a pas d'effet significatif sur la mesure de DTW ($b^{dtw} = 0.06\ [-0.00 - 0.12] 95\% IC$).

A noter aussi que pour les valeurs DTW les plus élevées correspondant au cluster 2, le passage des conditions de cluster 0 avec des retours audios vers le cluster 2 sans retours audios montre une augmentation légère de DTW ($b^{dtw} = 0.12\ [0.04 - 0.20] 95\% IC$).

Ainsi, une possibilité d'interprétation est que l'effet de la largeur/hauteur des images entraînent plus de dissimilarités entre les parcours visuels que l'absence ou la présence de retours audio.

TABLE 4.2 – Résultats issu du modèle bayésien pour étudier l'influence des conditions sur la similarité des observations par différents participants par la mesure DTW. La valeur de référence des modèles présentés sont des images appartenant au cluster 0 avec l'accompagnement de retours audios.

	Condition son : Cluster	Estimation	IC [95%]
DTW	Origine à l'ordonnée (Audio :cluster 0)	6.14	6.03 - 6.25
	SansAudio :cluster 0	0.06	-0.00 - 0.12
	Audio :cluster 1	0.60	0.47 - 0.73
	Audio :cluster 2	1.19	1.04 - 1.33
	SansAudio :cluster 1	0.06	-0.01 - 0.13
	SansAudio :cluster 2	0.12	0.04 - 0.20

B.2.2 Multimatch Le calcul des sous-métriques de Multimatch a été réalisé en comparant les parcours visuels des participants ayant regardé la même image dans les mêmes conditions.

La Fig. 4.16 récapitule les résultats obtenus sous la forme de boites à moustache et la dispersion des valeurs, à partir du calcul des métriques de similarité de Multimatch. L'interprétation des analyses bayésiennes s'appuie sur une référence où les images appartiennent au cluster 0 et la condition de son est fixée à la présence d'audio. Le tableau 4.3 récapitule les résultats pour les différentes sous-métriques de la métrique Multimatch suite à l'analyse bayésienne. Pour cette analyse la formule suivante a été utilisée pour chaque métrique :

$$MM^* = condition_{son} * cluster_{format} + (1|Participant_1) + (1|Participant_2) + (1|Image) \quad (4.4)$$

Il s'agit de la même formule que pour la modélisation de la mesure DTW.

Pour l'ensemble des sous-métriques multimatch calculés dans les différentes conditions, il y a très peu de différences qui sont marquées.

On peut cependant noter que les sous métriques MM_Len et MM_Shape ont des valeurs très élevées (puisque proche de 1) pour l'ensemble des conditions très peu de différences entre les conditions audios et les conditions de format.

La sous métrique MM_Pos montre aussi des valeurs élevées, mais principalement pour le cluster 0 ($M_{SA0}^{MM-Pos} = 0.95 \pm 0.02$)et le cluster 1 ($M_{SA1}^{MM-Pos} = 0.91 \pm 0.04$; $M_{A1}^{MM-Pos} = 0.92 \pm 0.03$). Le cluster 0 ne présente pas de différence avec ou sans retour audio, et cluster 2 présente très

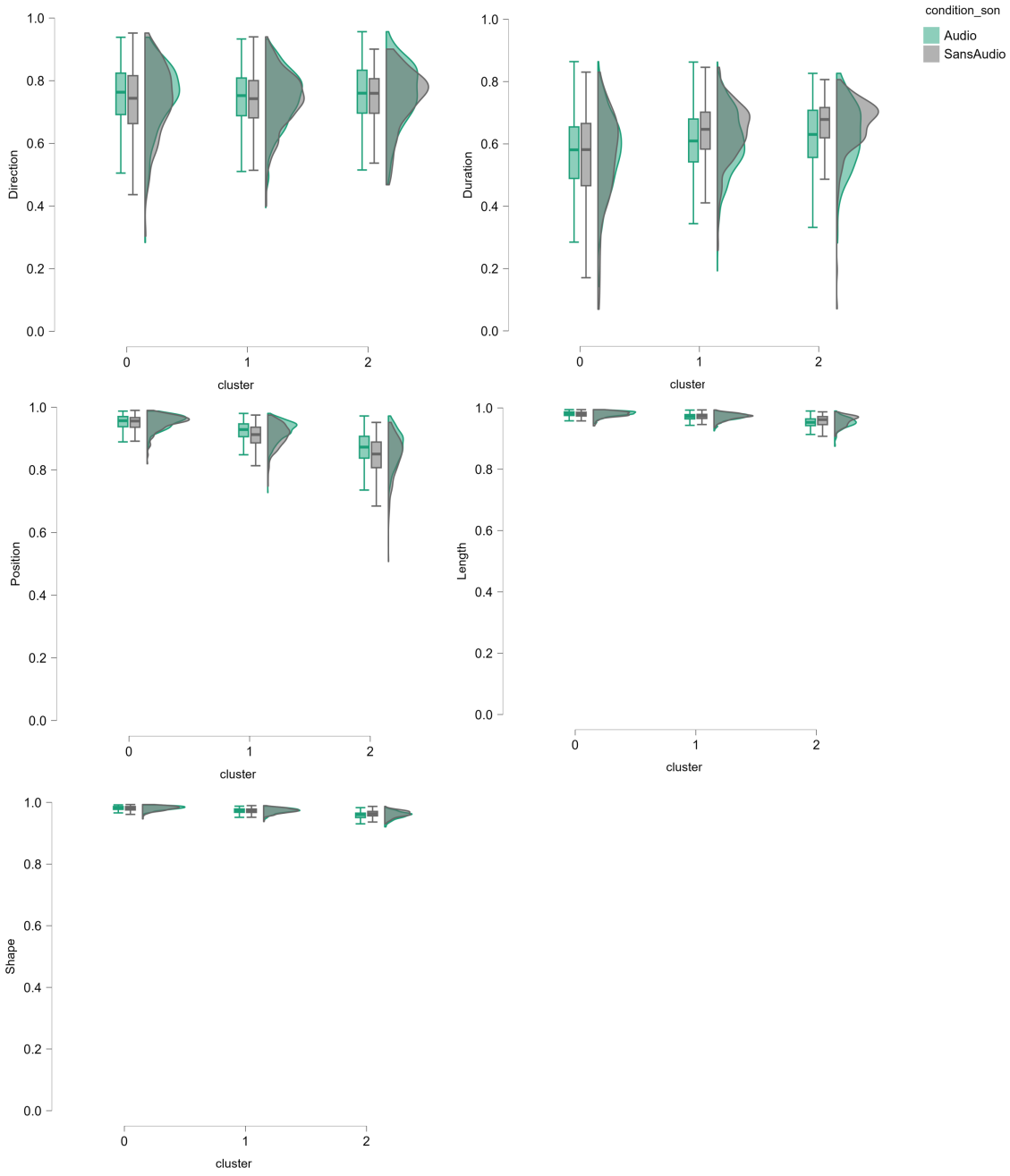


FIGURE 4.16 – Boîtes à moustache des résultats pour la métrique Multimatch

peu de différences entre les conditions audios. Pour cette sous-métrique MM_Pos, les valeurs du cluster 2 sont légèrement plus faibles ($M_{SA2}^{MM-Pos} = 0.84 \pm 0.07$; $M_{A2}^{MM-Pos} = 0.87 \pm 0.05$).

Pour les sous-métriques MM_Pos, MM_Len et MM_Shape, l'analyse bayésienne démontre, en restant sur la référence d'images observées avec de l'audio, que plus les images sont grandes, moins la forme (e.g, $b^{MM-Shape} = -0.80$ $[-0.91 - (-0.68)]95\%IC$), la longueur des vecteurs (cluster 2 :$b^{MM-Len} = -0.83$ $[-0.96 - (-0.69)]95\%IC$), ainsi que la position des points de fixations (cluster 2 :$b^{MM-Pos} = -1.04$ $[-1.19 - (-0.88)]95\%IC$) sont similaires.

Dans le cadre de la métrique MM_Dir, les résultats sont moyens, mais restent comparables pour les différents clusters avec la présence d'audio, ou son absence. L'analyse bayésienne avance que le changement de présence de retour audio à son absence pour des images appartenant toujours au cluster 0 entraîne une diminution assez légère de la métrique MM_Dir ($b^{MM-Dir} = -0.13$ $[-0.23 - (-0.02)]95\%IC$) et MM_Pos ($b^{MM-Pos} = -0.15$ $[-0.25 - (-0.04)]95\%IC$). Enfin, on observe des résultats moyens pour la sous-métrique MM_Dur (e.g., cluster 0 : $M_{SA0}^{MM-Dur} = 0.57 \pm 0.12$; $M_{A0}^{MM-Dur} = 0.55 \pm 0.15$). Avec une similarité des durées pour chaque point de fixation qui augmente légèrement lorsque l'on est toujours dans des conditions avec audio et que l'on passe à des images plus grandes : le cluster 1 $b^{MM-Dur} = 0.15$ $[0.05 - 0.26]95\%IC$ et cluster 2 $b^{MM-Dur} = 0.25$ $[0.13 - 0.36]95\%IC$.

Les résultats indiquent que la présence d'audio exerce un effet faible sur les métriques de similarité du regard (MM_Dur, MM_Dir, MM_Pos), avec des variations modestes et des intervalles de crédibilité proches de zéro. En revanche, le format des images influence de manière beaucoup plus marquée la similarité des trajectoires de regard, comme le montrent les valeurs négatives importantes des coefficients b pour MM_Shape, MM_Len et MM_Pos, en particulier pour le cluster 2. Ces observations suggèrent donc que la taille des images modifie davantage le comportement oculaire que la présence ou l'absence de retours audio dans le cadre de l'expérimentation menée.

TABLE 4.3 – Résultats issus des modèles bayésiens pour étudier l'influence des conditions sur la similarité des observations par différents participants par la métrique Multimatch. La valeur de référence des modèles présentés sont des images appartenant au cluster 0 accompagnées d'audio.

	Condition son : Cluster	Estimation	IC[95%]
MM-Dir	Origine à l'ordonnée (Audio :cluster 0)	1.12	1.01 - 1.23
	SansAudio :cluster 0	-0.13	-0.23 (-0.02)
	Audio :cluster 1	-0.04	-0.16 - 0.08
	Audio :cluster 2	0.00	-0.13 - 0.14
	SansAudio :cluster 1	0.05	-0.07 - 0.17
	SansAudio :cluster 2	0.04	-0.10 - 0.18
MM-Dur	Origine à l'ordonnée (Audio :cluster 0)	0.29	0.18-0.39
	SansAudio :cluster 0	-0.06	-0.16-0.04
	Audio :cluster 1	0.15	0.05-0.26
	Audio :cluster 2	0.25	0.13-0.36
	SansAudio :cluster 1	0.10	-0.01 - 0.21
	SansAudio :cluster 2	0.10	0.07- (-0.03)
MM-Pos	Origine à l'ordonnée (Audio :cluster 0)	2.97	2.84-3.10
	SansAudio :cluster 0	-0.15	-0.25-(-0.04)
	Audio :cluster 1	-0.44	-0.58 - (-0.30)
	Audio :cluster 2	-1.04	-1.19 - (-0.88)
	SansAudio :cluster 1	-0.06	-0.17-0.06
	SansAudio :cluster 2	-0.09	-0.20 - 0.03
MM-Len	Origine à l'ordonnée (Audio :cluster 0)	3.84	3.73-3.95
	SansAudio :cluster 0	-0.04	-0.13-0.04
	Audio :cluster 1	-0.32	-0.44-(-0.20)
	Audio :cluster 2	-0.83	-0.96 - (-0.69)
	SansAudio :cluster 1	0.08	-0.01-0.18
	SansAudio :cluster 2	0.17	0.07-0.27
MM-Shape	Origine à l'ordonnée (Audio :cluster 0)	3.93	3.83-4.03
	SansAudio :cluster 0	-0.04	-0.11-0.02
	Audio :cluster 1	-0.34	-0.44-(-0.24)
	Audio :cluster 2	-0.80	-0.91-(-0.68)
	SansAudio :cluster 1	0.05	-0.02-(0.11)
	SansAudio :cluster 2	0.16	0.09-0.22

C Retours sur l'expérience

La figure 4.17 présente les questions posées lors de l'entretien, à l'issue de l'observation des images par les participants, ainsi que la répartition en pourcentage des réponses sur une échelle de cinq points. Pour chaque énoncé, les participants devaient attribuer une note allant de 1 à 5 : 1 = "Pas du tout d'accord", 2 = "Pas d'accord", 3 = "Neutre", 4 = "D'accord", et 5 = "Tout à fait d'accord".

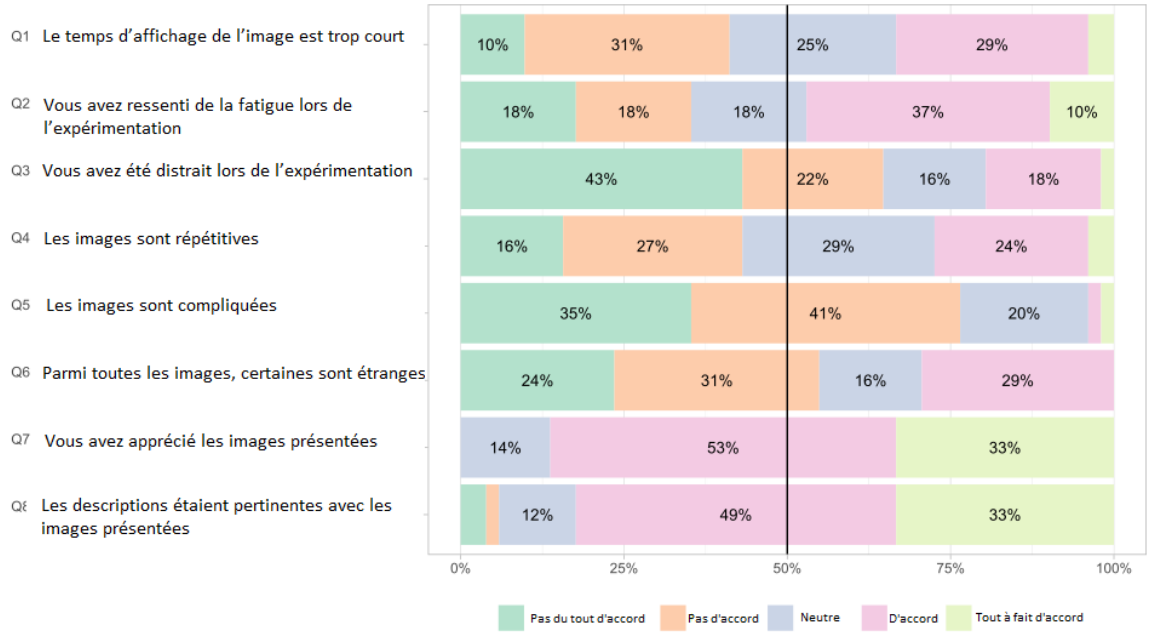


FIGURE 4.17 – Résultats pour l'échelle de Likert dans le cadre des retours de l'étude sur l'observation des images de la Tapisserie de Bayeux.

En ce qui concerne les ressentis sur les stimuli présentés, la durée d'affichage des images (6 secondes) a semblé légèrement trop courte pour les participants (Q1), et les images semblaient un peu répétitives (Q4), mais pas particulièrement compliquées (Q5). Les participants n'ont pas ressenti que la présentation des stimuli était perturbante (Q6). Néanmoins, la plupart des participants ont apprécié les images de la Tapisserie de Bayeux malgré leurs différentes présentations (Q7). En général, les courtes descriptions audio fournies pour certaines des images semblaient pertinentes (Q8). En termes d'expérience, certains participants se sont légèrement sentis fatigués pendant l'expérience (Q2), mais ne se sont pas sentis distraits (Q3).

3.4 Discussion

L'étude sur la saillance visuelle a montré que l'attention des participants se portait prioritairement sur les personnages, en particulier sur les mains et les visages. Ces éléments révèlent une importance particulière dans la Tapisserie de Bayeux, car ils jouent un rôle central dans la compréhension des scènes représentées. Par exemple, le geste du roi pointant du doigt un autre personnage peut être interprété comme une indication d'ordre, tandis que l'orientation des visages permet de déterminer l'interlocuteur des personnages ou l'objet de leur attention (e.g., la direction de leur déplacement).

En complément, les autres éléments ont été classés par ordre de priorité attentionnelle : d'abord les monuments, puis les objets, les animaux etc. Ces observations offrent des pistes pour améliorer la stratégie d'exploration hiérarchique. Dans le cadre de la théorie du gain de la cécité [12], il est proposé que l'exploration guidée globale s'appuie sur les indices de bas niveau (formes, couleurs, textures) décrites par une personne partiellement aveugle. L'exploration guidée détaillée, quant à elle, devrait mettre en valeur les éléments saillants selon leur importance attentionnelle, tout en intégrant les relations entre ces éléments.

Pour permettre l'accès à différents niveaux d'information, la F2T devrait offrir la possibilité de zoomer sur des zones précises des scènes. Par ailleurs, l'étude portant sur la similarité des parcours visuels a permis d'évaluer l'influence de la présence ou non de retours audio, ainsi que celle du format des images (petites, moyennes, grandes). Les résultats montrent que l'impact du format de l'image est plus déterminant que celui des retours audio. Cette observation pourrait s'expliquer par la nécessité, pour les participants, de partager leur attention entre l'image et la description verbale, ou encore par la nature relativement générale des descriptions, limitées à une durée de 6 secondes.

Ainsi, le séquencement de la présentation des éléments sur la F2T devrait s'appuyer sur les stratégies d'exploration visuelle observées. On pourrait favoriser des parcours haptiques principalement horizontaux pour les images larges et verticaux pour les images de format allongé. Ces recommandations constituent des pistes prometteuses pour guider l'exploration au moyen de retours haptiques.

En perspective, d'autres modalités pourraient être envisagées, telles que des explorations guidées par retour audio, ou encore l'intégration de modèles de VQA (Visual Question Answering), permettant aux utilisateurs de poser directement des questions sur le contenu de l'image. Des travaux connexes sur l'exploration de pages web par des personnes aveugles [192-194] offrent des solutions complémentaires. Le système TagThunder, par exemple, s'appuie sur une segmentation hiérarchique des pages web et l'identification de mots-clés, couplée à une spatialisation audio par effet cocktail, afin de permettre une exploration libre. Une première intégration expérimentale de TagThunder avec le logiciel F2T a été réalisée, aboutissant à un prototype nommé 2F4T, dont la faisabilité reste à évaluer.

Une question cruciale demeure : comment les musées pourraient-ils adapter ces dispositifs à d'autres types d'œuvres d'art ? Une piste de généralisation consisterait à recourir à des modèles d'apprentissage automatique. Dans le cas de la Tapisserie de Bayeux, une œuvre de grande ampleur, l'objectif serait de générer automatiquement des cartes de saillance et des parcours visuels pour des scènes qui n'ont pas fait l'objet d'analyses préalables. La section suivante propose ainsi d'étudier différents modèles capables de produire ces cartes et parcours, afin de les appliquer à d'autres scènes de la Tapisserie et, plus largement la méthodologie proposée pourraient êtres utilisées pour d'autres œuvres.

4 Sélection de modèles

4.1 Introduction

Afin d'automatiser l'adaptation des œuvres pour une exploration haptique via le dispositif F2T, il est nécessaire de chercher des modèles d'apprentissage automatiques capables de prédire de manière fiable les zones saillantes et les parcours visuels. La sélection de modèles adaptés constitue donc une étape clé : elle donne la possibilité de généraliser les observations issues d'un nombre limité d'expériences oculométriques à de nouvelles scènes non observées, et de réduire ainsi la dépendance à un recueil systématique de données par eye-tracker.

Le jeu de données utilisé dans cette étude comprend 314 images explorées librement par les participants, sans descriptions verbales. Ce corpus permet d'établir une vérité terrain de référence pour comparer différentes méthodes de prédictions de cartes de saillance et de trajectoires visuelles en exploration visuelle libre. Pour sélectionner et comparer efficacement des modèles capables de prédire les zones saillantes et les parcours visuels, il est nécessaire de rappeler les principales architectures de deep learning utilisées dans ce domaine. Les réseaux de neurones convolutifs (CNN), les LSTM (Long Short-Term Memory) et les Transformers représentent trois approches, chacune offrant des mécanismes distincts pour extraire, mémoriser ou relier des informations à partir des données en entrée. Dans ce cadre 3 modèles de prédictions de zones saillantes ont été étudiés : le modèle Resnet [155], modèle TranSalNet [195], et le modèle SAM-LSTM-Resnet [196]. Puis les modèles les plus performants en inférence sur la base de données de la Tapisserie de Bayeux ont été affinés avec ces mêmes données. De la même manière, trois modèles de prédiction de parcours visuels ont été choisis : le modèle bio-inspiré GEYMOL [197], le modèle de Chen *et al.* [198], et le modèle SPG-en [199].

L'objectif est d'identifier le modèle qui reproduit le plus fidèlement ces observations, afin de l'entraîner ensuite sur les données collectées et de l'intégrer dans un processus d'adaptation automatisé et inclusif.

4.2 Comparaison des méthodes de génération de cartes de saillance

A Méthodes & métriques

Dans la littérature, les chercheurs ont travaillé sur des modèles d'attention, en se concentrant sur les méthodes d'apprentissage automatique basées sur des réseaux de neurones convolutifs (CNN). Les CNN [200] sont des modèles capables d'extraire automatiquement des "features" (motifs locaux) dans les images, comme des bords ou des textures, grâce à des filtres de convolution appliqués sur l'image d'entrée. Ces motifs sont ensuite combinés hiérarchiquement dans des couches profondes pour représenter des structures de plus en plus complexes. Dans l'apprentissage supervisé, le modèle tente de découvrir ces structures à partir de données annotées. Les informations apprises peuvent ensuite être utilisées pour prédire l'attention visuelle sur de nouvelles images.

Pour la prédiction des cartes de saillance basée sur les fixations oculaires humaines, Cornia *et al.*

[196] ont proposé un modèle d'attention basé sur la saillance utilisant un réseau LSTM . Les LSTM sont des réseaux de neurones récurrents (RNN) conçus pour mémoriser des informations sur de longues séquences, ce qui leur permet de capturer des dépendances temporelles dans des données séquentielles [201]. Lou *et al.* [195] ont incorporé des Transformers pour améliorer les performances de la prédiction de la saillance. Les Transformers sont des modèles basés sur l'attention qui utilisent un encodeur pour analyser les entrées et un décodeur pour générer des sorties. Ils permettent de récupérer les relations entre éléments éloignés entre eux, contrairement au traitement séquentiel réalisé par les RNNs. ResNet50 [155] est une architecture de type CNN composée de 50 couches de réseau de neurones destinée à détecter les régions saillantes dans les images. Ici, il est utilisé comme base dans SAM-LSTM-ResNet et TranSalNet.

L'architecture du modèle de Cornia *et al.* [196], nommé SAM-LSTM-ResNet, pour Saliency Attentive Model-LSTM-ResNet, est présenté dans la figure 4.18. Elle comprend quatre modules clés : convLSTM (Convolutional LSTM) attentif ("Attentive convLSTM"), probabilités a priori apprises ("learned priors"), un CNN dilaté et la fonction de perte. ConvLSTM est une variante du modèle LSTM qui intègre des filtres de convolutions. Il permet ainsi de traiter des informations à la fois temporelles et spatiales. Ici, il est utilisé pour analyser et retenir les caractéristiques de l'image au fil du temps pour générer des cartes de saillance. Les probabilités a priori apprises ("learned priors") correspondent à la modélisation du biais central à l'aide de fonctions gaussiennes.

Le CNN dilaté correspond à l'intégration d'opérations de convolution dilaté à un CNN. La convolution dilaté permet d'utiliser un champ perceptif plus large que le simple CNN en jouant avec un paramètre nommé taux de dilatation. Ainsi, en plus d'avoir accès aux détails fins de l'image, il permet de prendre en compte un contexte plus large.

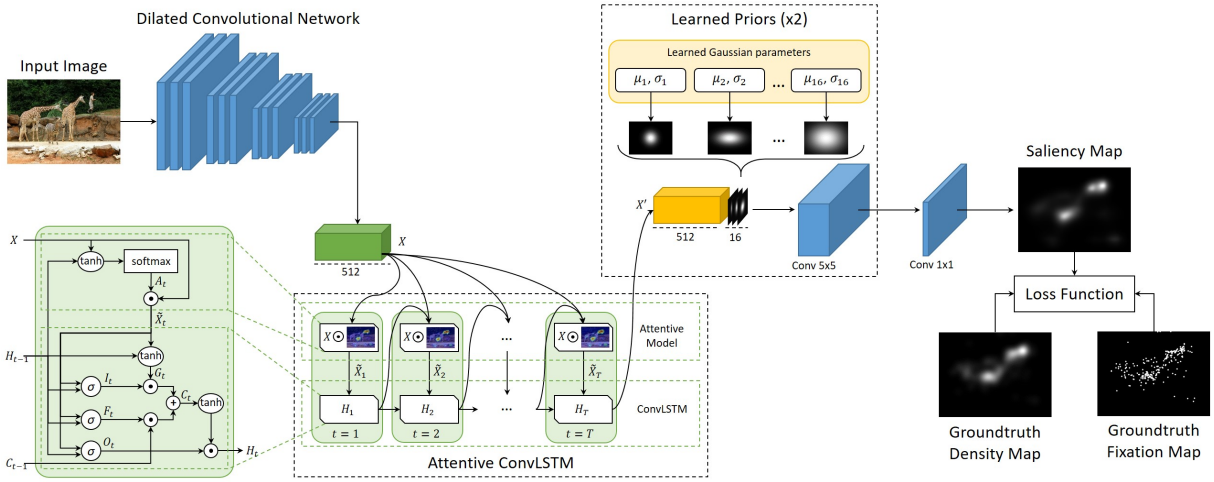


FIGURE 4.18 – Architecture de SAM-LSTM-ResNet[196]

L'architecture de TranSalNet [195], présentée dans la figure 4.19, combine des couches CNN avec des encodeurs Transformer pour la prédiction de la saillance. L'encodeur CNN, basé sur ResNet50, extrait trois cartes de caractéristiques à différentes échelles. Chacune de ces cartes est traitée par un encodeur Transformer séparé afin d'incorporer des informations contextuelles à longue portée. Le décodeur CNN fusionne ensuite ces cartes enrichies pour produire une carte de saillance finale à la résolution de l'image, en utilisant des opérations de sur-échantillonnage ("upsampling") pour restaurer la taille originale et des sauts de connexions ("skip connections")

pour combiner les informations locales de l'encodeur avec les cartes traitées par le décodeur, améliorant ainsi la précision spatiale.

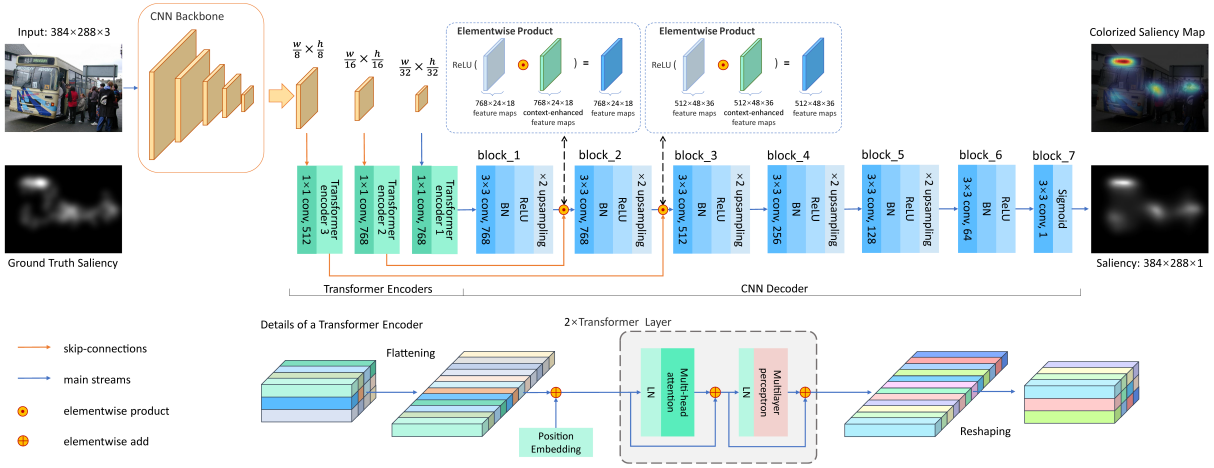


FIGURE 4.19 – Architecture TranSalNet-ResNet [195]

Les deux modèles TranSalNet et SAM-LSTM-ResNet utilisent la même fonction de perte, calculée à l'aide de trois métriques d'évaluation : NSS, KL-Div et CC.

— La métrique NSS (Normalized Scan path Saliency) est spécifiquement conçue pour l'évaluation des modèles de saillance. Elle quantifie les valeurs de saillance aux emplacements de fixation et les normalise en utilisant la variance de la carte de saillance.

— La métrique KL-Div (Kullback-Leibler Divergence) mesure la différence statistique entre la carte de saillance prédite et la distribution de densité de la vérité terrain.

— La métrique CC (coefficient de corrélation de Pearson) évalue la perte d'information entre les cartes de saillance prédites après seuil et la vérité terrain.

SAM-LSTM-ResNet [196] et TranSalNet [195] ont été entraînés sur le jeu de données Salicon [178], tandis que ResNet50 [155]a été entraîné sur le jeu de données ImageNet [202].

Pour l'application des prédictions de ces modèles dans l'adaptation des images de la tapisserie de Bayeux pour un dispositif haptique, les modèles ont été entraînés sur le jeu de données présenté dans la section 3.

Ces modèles ont été adaptés au jeu de données de la Tapisserie de Bayeux afin de déterminer quel modèle prédit le mieux les données collectées avec la plus grande similarité. Le jeu de données pour l'entraînement est composé de 308 images observées librement sans description verbales issues de la collecte de données. Il est réparti en 70% pour l'entraînement, 15% pour la validation et 15 % pour les tests.

Étant donné la complexité élevée de ces modèles par rapport à la taille limitée du jeu de données de la tapisserie de Bayeux, quelques modifications ont été nécessaires. Pour SAM-LSTM-ResNet [196], des couches de dropout et de normalisation ont été ajoutées pour améliorer la généralisation. Dans le cas de TranSalNet [195], les couches supérieures ont été gelées pour conserver les représentations de caractéristiques de haut niveau du modèle pré-entraîné, tandis que seules les couches 4,5,6 du décodeur CNN ont été autorisées à être entraînées sur le jeu de données, tout en introduisant des couches de dropout pour éviter le sur-apprentissage. De plus, une technique d'arrêt précoce basée sur les performances de validation a été appliquée pour cette même raison.

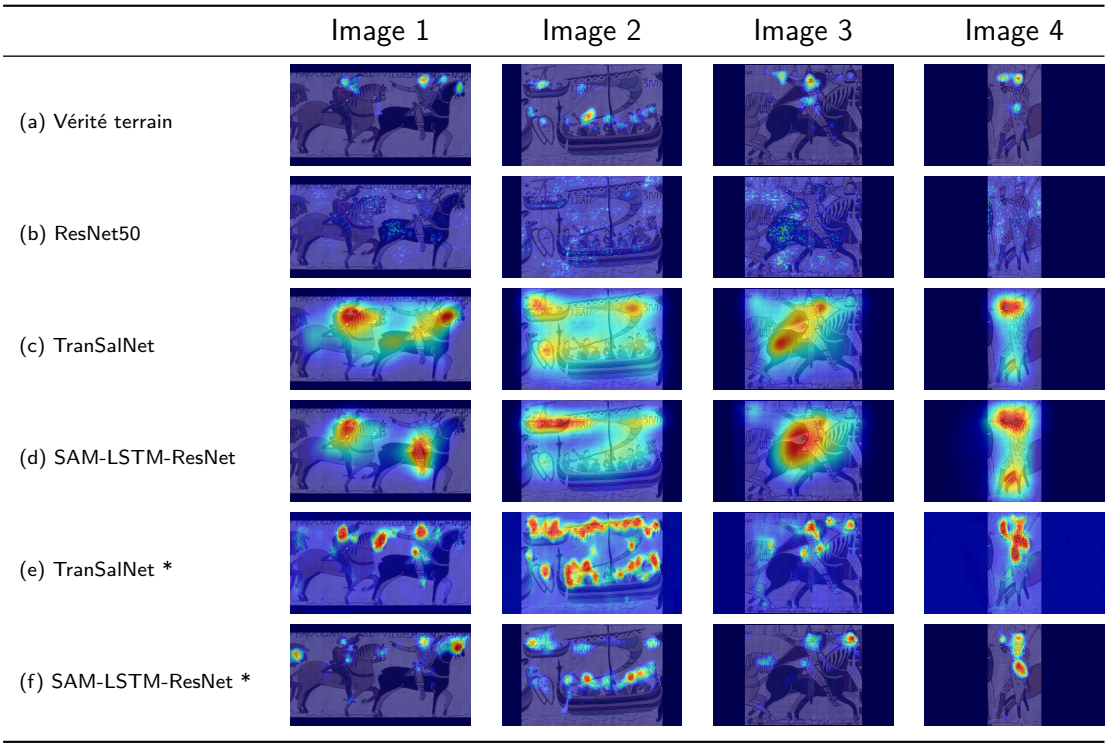


FIGURE 4.20 – Tableau de comparaison des cartes de saillance pour différents modèles la vérité terrrain (a), ResNet50 (b), TranSalNet(c), SAM-LSTM-ResNet (d), TranSalNet entraîné (e), SAM-LSTM-ResNet entraîné (f). (*) entraîné sur la base de données de la Tapisserie de Bayeux

B Étude qualitative

La figure 4.20 présente un tableau comparatif des prédictions réalisées par les modèles décrits ci-dessus par rapport à la vérité terrain.

Les cartes de saillance de la vérité terrain, moyennées à partir des données de suivi oculaire de plusieurs participants (Fig. 4.20.a), révèlent des régions saillantes localisées, principalement centrées sur les visages des personnages, les chevaux et les mains.

Le modèle ResNet50 [155] (Fig. 4.20.b) ne détecte aucune caractéristique distinctive dans la scène. En revanche, les modèles SAM-LSTM-ResNet [196] Fig. 4.20.c) et TranSalNet [195] (Fig. 4.20.d) mettent en évidence des régions saillantes plus larges, correspondant aux éléments saillants observés dans la vérité terrain, en plus de capturer des éléments textuels.

Après entraînement sur la base de données de la Tapisserie de Bayeux, TranSalNet et SAM-LSTM-ResNet montrent une amélioration significative de la précision des prédictions. Comme observé dans les images 2 et 4 de la figure 4.20, les régions de saillance prédites par TranSalNet (Fig.4.20.e) restent assez larges par rapport à SAM-LSTM-ResNet (Fig. 4.20.f). TranSalNet prédit beaucoup plus de régions saillantes en comparaison avec la vérité terrain (Fig. 4.20.a). En revanche, lorsqu'on observe les résultats issus de l'entraînement avec SAM-LSTM-ResNet, les régions de saillance prédites sont beaucoup plus alignées avec la vérité terrain, avec des limites qui correspondent étroitement à celles observées dans la vérité terrain(Fig. 4.20.f).

C Étude quantitative

Pour comparer ces modèles de manière quantitative, les métriques AUC Judd et SIM (similarité), couramment utilisées dans la littérature, ont été calculées sur les images du jeu de test. La table 4.4 indique les moyennes et les écarts-types des modèles testés.

Le modèle TranSalNet, entraîné sur les données de la Tapisserie de Bayeux, a obtenu les meilleurs résultats pour l'AUC (Area under ROC Curve), tandis que le modèle SAM-LSTM-ResNet a obtenu de meilleurs résultats pour la SIM. Étant donné que les prédictions du modèle SAM-LSTM-ResNet sont plus similaires aux données collectées, ce modèle peut être sélectionné comme modèle choisi pour les utilisations futures.

TABLE 4.4 – Métriques de saillance pour les modèles testés

Modèles	AUC↑	SIM↑
ResNet50	$0.888[0.102]$	$0.297[0,088]$
SAM-LSTM-ResNet	$0.891[0.098]$	$0.291[0,083]$
TranSalNet	$0.908[0.076]$	$0.313[0.078]$
SAM-LSTM-ResNet*	$0.895[0.091]$	$0.355[0.157]$
TranSalNet*	$0.917[0.067]$	$0.324[0.072]$

4.3 Comparaison des méthodes de génération de parcours visuels

A Méthodes

Les prédictions des parcours de fixation sont générées ici en utilisant trois modèles [197-199] et évaluées sur 314 images de la tapisserie observées par les participants lors de la collecte des données.

Le modèle de Chen *et al.* [198] et le modèle de Kerkouri *et al.* : SPG-En [199] génèrent des parcours de fixation visuelle pour des tâches de vision libre à l'aide de modèles d'apprentissage profond, tandis que G-Eymol [197] est un modèle bio-inspiré basé sur les lois gravitationnelles. Pour les images statiques, G-Eymol s'appuie sur des caractéristiques d'image de bas niveau et intègre le mécanisme d'inhibition du retour pour générer des parcours de fixation.

Le modèle proposé par Chen *et al.* [198], dont l'architecture est présenté dans la figure 4.21, est une approche d'apprentissage par renforcement profond, entraînée à l'aide de l'entraînement séquentiel auto-critique (SCST - Self-Critical Sequence Training). Il est conçu pour plusieurs tâches, notamment la réponse à des questions visuelles (VQA - Visual Question Answering), la recherche visuelle et la vision libre. Cette étude se concentre sur l'aspect vision libre, qui a été formé sur le jeu de données OSIE. Son architecture comprend un module d'attention et un réseau ConvLSTM, similaire à l'approche proposée par Cornia *et al.* [196]. Une caractéristique clé du ConvLSTM de Chen *et al.* est sa capacité à prédire la position, l'ordre et la durée des fixations.

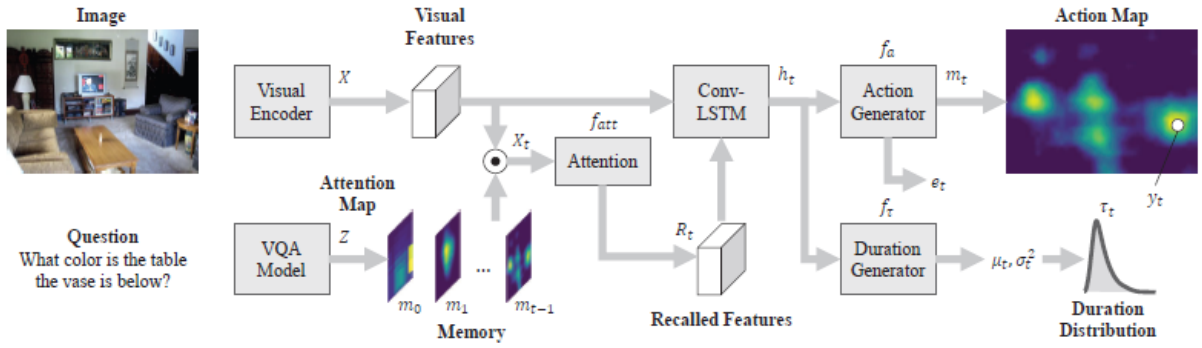


FIGURE 4.21 – Architecture du modèle de Chen [198]

SPG-En[199], dont l'architecture est présenté dans la figure 4.22, est un modèle léger pour la prédiction de parcours de fixation, formé sur le jeu de données SALICON. Il utilise une architecture MobileNet [203] pour extraire les caractéristiques des images et emploie un décodeur pour fusionner les caractéristiques extraites avec des cartes basées sur des distributions gaussiennes qui simulent le biais central lors de l'observation d'image. De plus, il intègre une couche de sélection, agissant comme un masque pour déterminer si des points de fixation spécifiques font partie de la séquence prédite. De plus, SPG-En utilise un algorithme d'adaptation de domaine non supervisé et adversarial pour ajuster les prédictions pour différents domaines d'images, spécifiquement pour passer des images de scènes naturelles aux peintures. Les auteurs visaient à prédire les trajectoires de fixation sur des peintures ; cependant, en raison du manque de jeux de données contenant des informations

sur les fixations des yeux (fixations, saccades) pour les peintures, le modèle n'a pas pu être pleinement évalué dans ce contexte.

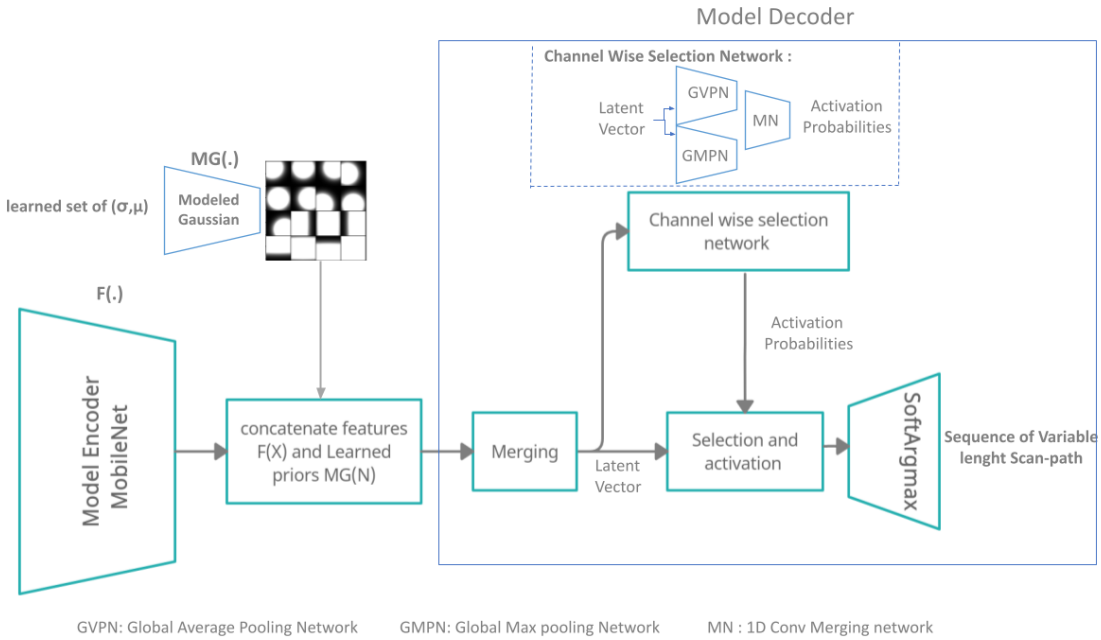


FIGURE 4.22 – Architecture de SPGen[199]

Pour évaluer les performances des modèles sur l'inférence de ces modèles sur les images de la Tapisserie de Bayeux, les métriques Dynamic Time Warping (DTW) [189] et MultiMatch [190] ont été calculées.

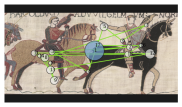
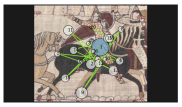
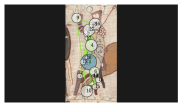
B Etude qualitative

Cette section explique les parcours de fixation prédits par les modèles présentés ci-dessus sur l'image de la Tapisserie de Bayeux. La Figure 4.5 illustre les trajectoires visuelles prédites par ces modèles, offrant une comparaison directe entre eux.

Figure 4.5.a présente le parcours de fixation de vérité terrain d'un participant. Comme observé dans la section précédente, la vérité terrain montre que les fixations sont principalement localisées sur les têtes des personnages, les mains et les éléments textuels (Image 1 de la Fig. 4.5.a). Il est notable que les saccades apparaissent plus courtes sur les écritures (Image 1) et plus longues lors des transitions entre les éléments distincts (par exemple, du navire au premier plan vers le petit bateau dans le coin supérieur gauche de l'Image 2 de la Fig. 4.5.a). Les angles entre les saccades sont principalement horizontaux.

Pour la méthode G-Eymol, une durée de 6 secondes a été définie pour correspondre aux conditions de l'expérience. Cette approche bio-inspirée conduit à un nombre de saccades significativement plus élevé par rapport à la vérité terrain. Le modèle met en évidence un biais central et de nombreuses saccades, influencées en grande partie par le mécanisme d'inhibition du retour. En conséquence, on observe que les saccades sont systématiquement plus longues lors des transitions entre les éléments saillants (Image 2, Fig. 4.5.b), et les fixations restent principalement centrées

TABLE 4.5 – Tableau de comparaison de parcours visuels générés à partir d'images issues de la Tapisserie de Bayeux avec la vérité terrain (a) le modèle G-EYMOL [197] (b), le modèle de Chen [198] (c) et le modèle SPGen de [199] (d)

	Image 1	Image 2	Image 3	Image 4
(a) Vérité terrain				
(b) G-EYMOL				
(c) CHEN				
(d) SPGen				

au milieu de l'image.

En revanche, le modèle de Chen se concentre sur l'identification d'éléments saillants plus contextuels, en particulier les visages (Image 2, Fig. 4.5.d), et génère des saccades plus longues. Ce modèle simule également le mouvement de va-et-vient des regards entre les éléments clés, imitant le comportement du participant.

Le modèle SPGen, quant à lui, ne priorise pas des éléments spécifiques comme les visages ou les mains. Au lieu de cela, il met l'accent sur les relations entre les éléments de la scène. Il met en évidence les parties saillantes de l'image en fonction du centre de masse, plutôt qu'en se basant sur des caractéristiques sémantiques telles que les visages ou les mains.

C Etude quantitative

Pour comparer quantitativement ces modèles, les résultats des calculs de DTW (Dynamic Time Warping) et Multimatch sont présentés dans le tableau 4.6. Dans l'ensemble, le modèle de Chen produit les inférences de parcours de fixation les plus précises pour le jeu de données de la Tapisserie de Bayeux. Il donne les meilleurs résultats pour les positions des fixations et la longueur des vecteurs de saccade, ce qui en fait un modèle fiable pour ce jeu de données, et présente une dissimilarité moindre que les autres modèles.

G-Eymol, bien que moins précis en termes de placement des fixations, excelle dans la prédiction de la précision directionnelle. Bien que SPG-En semble moins adapté pour ce type d'œuvre d'art, il parvient bien à capturer la forme des trajectoires inférées et, tout comme Chen, il est performant en termes de longueur des vecteurs de saccades, fournissant des informations sur la structure globale du mouvement du regard.

Ainsi, sur la base de cette comparaison, le modèle de Chen émerge comme le modèle le plus approprié pour le jeu de données utilisé dans cette étude. De plus, sa capacité à prendre en compte

TABLE 4.6 – Métriques pour les parcours visuels pour les modèles testés

Modèles	DTW↓	MM-Dir↑	MM-Len↑	MM-Pos↑	MM-Shape↑	MM-Dur↑
G-Eymol	2161[386]	**0.686[0.087]**	0.940[0.010]	0.706[0.024]	0.959[0.004]	0.600[0.100]
Chen	**974[1132]**	0.655[0.121]	**0.977[0.013]**	**0.733[0.055]**	0.981[0.009]	**0.684[0.105]**
SPGen	3834[623]	0.548[0.116]	**0.977[0.018]**	0.557[0.091]	**0.985[0.009]**	

les durées de fixation le rend particulièrement précieux pour les applications d'adaptation d'images.

5 Conclusion du chapitre

L'objectif principal de cette recherche est d'analyser comment il est possible de présenter une œuvre d'art aux personnes aveugles et partiellement aveugles en utilisant des sens autres que la vision. Comme introduit dans la première section de ce chapitre, les personnes aveugles et partiellement aveugles ont exprimé leur préférence pour explorer une scène de manière hiérarchique, en partant d'une représentation globale pour arriver aux détails, comme démontré dans les tests précédents 3. L'exploration globale peut correspondre à un premier aperçu d'une scène, offrant ainsi une analyse structurée de l'œuvre. Le contexte de la découverte des œuvres est également crucial.

Briber *et al.* [204] rapportent que l'expérience artistique est perçue comme plus intéressante dans le cadre d'un musée que sur un ordinateur distant. Les visiteurs passent davantage de temps à analyser et à partager leur expérience dans le contexte réel. Walker *et al.* [205] ont analysé les modèles de mouvement oculaire d'enfants et d'adultes au Musée Vincent Van Gogh, concluant que les adultes, plus que les enfants, suivaient un schéma descendant lorsqu'ils observaient une peinture. Ils commençaient par une vision globale avant de se concentrer sur les détails, avec une attente et une intention particulières. Le regard des adultes était moins dispersé que celui des enfants.

Pour les personnes aveugles et partiellement aveugles, une représentation globale d'une œuvre d'art pourrait inclure des éléments saillants, ainsi que leur localisation dans la scène et leurs interactions avec les autres éléments. Il est essentiel d'éviter une surcharge cognitive et de suivre les régions saillantes ainsi que les parcours visuels observés dans ce travail de recherche, afin d'adapter le guidage de l'exploration haptique avec des descriptions audio.

Les questions de recherche sont posées afin de guider l'orientation des meilleures pratiques à partir des résultats obtenus.

1. Comment identifier les éléments sémantiquement importants dans une œuvre d'art ?

Pour identifier les éléments sémantiquement importants, une étude de suivi oculaire a été réalisée sur une œuvre d'art de taille considérable (700 cm x 50 cm). Cette étude a permis de générer des cartes de saillance et des représentations des parcours visuels pour un nombre limité d'images de la Tapisserie de Bayeux. Dans ce contexte, une comparaison a été effectuée avec les méthodes récentes basées sur la saillance, étudiées pour générer des cartes de saillance sur des scènes entières ou partielles de la Tapisserie de Bayeux. Compte tenu des particularités de l'œuvre et des conditions d'observation, les algorithmes les plus appropriés ont été sélectionnés pour transcrire les données collectées.

Les résultats ont mis en évidence que les participants ont tendance à observer les visages, les mains et le texte dans le contexte des images de la Tapisserie de Bayeux. Ces résultats sont cohérents avec les études comportementales de la littérature, telles que celles de Cerf *et al.*[183] et Langton *et al.* [206]. En effet, les participants privilégient ces éléments par rapport aux arbres ou aux éléments architecturaux. Un élément particulier de la Tapisserie de Bayeux réside dans les mains : les personnages brodés avec les mains levées sont interprétés comme ceux qui parlent ou

donnent des ordres. Les mains levées figurent fréquemment parmi les éléments saillants.

Il est donc proposé de commencer la représentation globale haptique d'une scène avec des zones de saillance prédites. Les jeux de données pour l'entraînement pourraient être choisis parmi les jeux de données disponibles sur des images naturelles, tels que Salicon, MIT1003 ou un jeu de données sur les peintures, tel que celui de Le Meur pour un premier apprentissage. Puis, un ajustement sur un jeu de données spécialisé pourrait être utilisé pour améliorer la qualité des résultats.

A partir du jeu de données constitué, des modèles de cartes de saillance et de prédiction de parcours de fixation ont été comparés afin d'obtenir une adaptation meilleure. Les modèles de Cornia *et al.* [196] et de Chen *et al.* [198] se sont avérés les plus efficaces pour généraliser la prédiction des cartes de saillance et des parcours visuels pour les images de la Tapisserie de Bayeux, respectivement.

2. Quelle est l'ordre d'importance des éléments scannés ?

Les observations des scènes et les résultats statistiques tendent à montrer qu'il existe un biais central dans les parcours visuels. Cela est conforme aux résultats de Bindemann [207], qui confirme qu'un biais central existe pour les mouvements oculaires, quel que soit l'image donnée. Ainsi, l'ordre de présentation des éléments saillants commence généralement près du centre de la scène. Une légère tendance à s'orienter d'abord vers la gauche, puis vers la droite, a également été observée. Comme la Tapisserie de Bayeux raconte une histoire avec des scènes allant de gauche à droite, ce mouvement pourrait être un comportement naturel propre à cette œuvre. Le texte en latin en haut des scènes constitue également une zone d'attention et suit également une direction de gauche à droite. L'eye tracker a enregistré de nombreuses trajectoires de balayage entre les éléments interactifs tels que les mains et les visages, ou entre les lettres d'un mot. Cela peut être reproduit avec des effets haptiques améliorés par des descriptions audio pour la dispositif haptique à retour de force, F2T.

3. L'audio-description influence-t-elle l'exploration de l'image ?

Dans cette étude, bien que les descriptions verbales aient été généralement considérées comme pertinentes pour les images présentées, elles étaient très courtes et pas suffisamment informatives sur le contexte. En ce qui concerne la saillance, une plus grande dispersion des zones saillantes a été observée sans descriptions verbales, par rapport à des zones plus concentrées avec descriptions. Cela est en accord avec les travaux de Krejtz *et al.* [208], qui a démontré le bénéfice des descriptions auditives pour orienter l'attention des enfants.

L'analyse statistique utilisant la mesure DTW et les métriques Multimatch indiquent que les descriptions verbales améliorent la concentration sur des éléments dispersés dans des images entières. En revanche, pour des formats plus petits, les descriptions verbales ont également concentré l'attention, mais cela était probablement dû à la plus petite taille de la zone visible. Il est intéressant de noter que les descriptions verbales ont eu moins d'impact que le format présenté dans cette étude.

Ces résultats suggèrent des considérations clés pour les adaptations proposées. Plus précisément, le système F2T pourrait intégrer un module de zoom avec une nouvelle image pour mettre en évidence les détails. Une approche potentielle serait de proposer séquentiellement les descrip-

tions verbales avant l'exploration afin de renforcer l'expérience tactile lors de l'interaction avec des éléments individuels.

Étant donné que les participants étaient des adultes, l'impact des descriptions verbales pourrait avoir été réduit. L'étude de Walker *et al.* [205] soutient cette idée, montrant des différences minimales dans l'exploration des adultes en utilisant des processus ascendant (basés sur le stimulus) ou descendant (basés sur le contexte). En revanche, les enfants ont tendance à explorer en utilisant un processus ascendant (basé sur le stimulus). Ainsi, les descriptions audio devraient être adaptées au profil de l'utilisateur. Pour la F2T, qui vise à être un dispositif inclusif pour les utilisateurs aveugles et partiellement aveugles, il est crucial d'adapter les descriptions afin d'éviter toute redondance, que ce soit par perception auditive ou haptique.

Certaines limitations peuvent néanmoins être notées pour cette étude. En effet, il n'a pas été possible d'étudier la perception des personnes partiellement aveugles. En raison des limites du suivi oculaire, par exemple le dispositif Eye-tracker a eu des difficultés à calibrer le regard avec deux yeux qui pointent dans deux directions différentes.

Cette étude se concentre sur la Tapisserie de Bayeux, il serait intéressant de comparer avec d'autres œuvres d'art, telle que la Tapisserie de l'Apocalypse qui est elle aussi une œuvre de style médiévale.

5 Conclusion

Sommaire

1 Conclusion de la thèse 109
2 Perspectives 110

1 Conclusion de la thèse

Cette thèse s'est inscrite dans la problématique de l'accessibilité des œuvres d'art pour les personnes aveugles et partiellement aveugles, en explorant les modalités sensorielles et technologiques permettant de proposer une expérience muséale riche au-delà de la seule vision. L'analyse des travaux existants a mis en évidence l'importance des informations sémiotiques (e.g. contours, formes, textures) ainsi que des informations sémantique (e.g., localisation, contexte) dans les processus de restitution multisensorielle. La littérature a montré que les représentations actuelles privilégient soit les dispositifs en relief, riches mais peu modulables, soit les interfaces numériques, flexibles mais souvent limitées à des retours unimodaux ou limité dans les informations transcrites. Ceci révèle un besoin de solutions combinant modularité, richesse perceptive et complémentarité des modalités sensorielles en passant par des dispositifs haptiques ne se basant pas sur des représentations tangibles.

Les études menées dans le cadre de cette recherche ont confirmé la pertinence d'une approche hiérarchisée, où l'exploration tactile et auditive progresse du global vers le détail. Les expérimentations ont également mis en lumière l'efficacité des dispositifs audio-haptiques pour structurer la compréhension d'une œuvre, en associant la perception des contours et volumes à des descriptions verbales contextualisées. Le prototype F2T, développé et évalué au cours de cette recherche, a montré qu'une interaction multimodale modulable pouvait constituer un support adapté aux attentes d'un public aveugle, partiellement aveugle et non aveugle, tout en répondant aux contraintes propres au milieu muséal. Un mode d'exploration hiérarchique a été proposé composé de 3 phases : exploration guidée globale, exploration guidée détaillée, exploration sans guidage.

Où l'exploration guidée globale est utilisée par des utilisateurs de manière passive pour obtenir des informations sémantiques sur l'œuvre. L'exploration guidée détaillée est décomposée en deux étapes pour permettre à la fois une exploration active et passive pour obtenir des détails plus fins sur les contours à travers le mouvement et les retours audios. Enfin une exploration sans guidage est proposée permettant d'exploiter des éléments de sonification de scènes et d'effets de frictions et de flux à travers des retours de force pour l'exploration indépendante des utilisateurs. Néanmoins, ces étapes d'exploration reste à être affinées, et plus particulièrement l'exploration sans guidage. Une des limites importante dans l'exploitation du dispositif F2T est l'évaluation de la perception

haptique individuelle des effets de frictions par retour de force, et notamment leur intégration multiple dans une même scène.

Un apport méthodologique supplémentaire a été l'intégration de techniques issues de la vision par ordinateur et de l'intelligence artificielle, utilisées notamment pour analyser les parcours visuels de spectateurs non aveugles et identifier des zones de saillance. Cette étape n'a pas pour objectif de centrer l'expérience sur la vision, mais de guider la hiérarchisation des informations et la génération de représentations audio-haptiques plus pertinentes, en mettant en avant des éléments saillants de manière multimodale. Ces résultats, combinés aux avancées en segmentation automatique, ouvrent la voie à une production semi-automatisée de contenus accessibles, pouvant contribuer à élargir l'accès à l'art.

Certaines limites doivent toutefois être soulignées. Les expérimentations se sont concentrées sur un corpus restreint d'œuvres, ce qui limite la généralisation des résultats, bien que l'objectif fût de comparer différents dispositifs pour une même œuvre. La participation réduite de personnes partiellement aveugles constitue également une restriction. Enfin, la surcharge cognitive induite par la combinaison de modalités sensorielles lors de l'exploration sans guidage demeure une problématique ouverte, appelant à une conception attentive à l'équilibre entre richesse perceptive et simplicité d'usage.

2 Perspectives

Les travaux réalisés ouvrent plusieurs pistes de prolongement. La première concerne l'automatisation des chaînes de traitement : les progrès rapides de l'apprentissage profond et de la vision par ordinateur permettent désormais de générer quasi automatiquement des représentations audio-haptiques à partir d'images d'œuvres d'art, tout en intégrant des mécanismes de hiérarchisation adaptés aux besoins perceptifs des utilisateurs.

Un second axe de recherche réside dans l'élargissement des corpus et des contextes étudiés. L'application des méthodes développées à d'autres œuvres et environnements muséaux permettrait de tester la robustesse des modèles et d'enrichir la diversité des expériences accessibles. L'exploration d'autres modalités sensorielles, telles que l'olfactif ou la vibration enrichie, pourrait également offrir de nouvelles formes d'immersion multisensorielle.

Enfin, une perspective plus large concerne la conception même des dispositifs muséaux. L'intégration de la dimension inclusive dès la phase de conception, plutôt que dans une logique d'adaptation a posteriori, constitue un horizon souhaitable. A terme, cette approche pourrait contribuer à redéfinir les propositions de la médiation culturelle, en plaçant les dispositifs haptiques au cœur de l'innovation muséale.

Publications

- L.Djoussouf, E. Pissaloux, N.T. Truong, K. Romeo, S. Gay, S. D. Dao « Vers des représentations multimodales de peintures pour leur accessibilité inclusive », Conf. Nat. Handicap 2022, 8-10 juin 2022, Paris (FRANCE).
- K. Romeo, E. Pissaloux, S. L. Gay, N.-T. Truong, et L. Djoussouf, « The MAPS : Toward a Novel Mobility Assistance System for Visually Impaired People », MDPI Sensors, vol. 22, no 9, Art. no 9, janv. 2022, doi :10.3390/s22093316.
- Pissaloux, E., Djoussouf, L., Romeo, K., Velazquez, R., Gay, S., Truong, N.G, Dao, S-D., « A new approach to assist Virtual Image accessibility for Visually Impaired People", HCI International 2022, Sweden, 2022
- L. Djoussouf, K. Romeo, « L'accessibilité des images pour les non-voyants : application au patrimoine culturel ». ORASIS 2023. Mai 2023, Carqueiranne(FRANCE).
- L. Djoussouf, K. Romeo, M. Chottin, H. Thompson, A.-F. Eardley, « Inclusion for Cultural Education in Museums, Audio and Touch Interaction », AAATE 2023, August 30-September 1st, 2023
- L. Djoussouf, K. Romeo, A. El Moataz, « Vers une approche inclusive de description multimédia pour le patrimoine», GRETSI 2023, Grenoble. Août-Septembre 2023 (poster)
- W. Luo, L. Djoussouf, C. Lecomte, K. Romeo, « Detecting areas of interest for Blind People », ICCHP 2024, 8-12 Juillet, Linz (Autriche)
- M. Redon, L. Djoussouf, K. Romeo, Y. Quéau, A. El Moataz Billah, « A Comparison Of Audio-Tactile Exploration Methods To Discover The Apocalypse Tapestry", ICCHP 2024, 8-12 Juillet, Linz (Autriche)
- L. Djoussouf, N. Charafe ,I. Mohammed-Riad Zizi , C. Lecomte, K.Romeo , Hierarchical exploration of artworks through 2-DOF haptic device : a usability study, WIP Eurohaptics 2024 (poster), Lille (FRANCE).

Bibliographie

[1] Polly Edman. “Production Methods and Materials”. In : *Tactile graphics*. American Foundation for the Blind, 1992.

[2] Richa Gupta, M. Balakrishnan et P.V.M. Rao. “Tactile Diagrams for the Visually Impaired”. In : *IEEE Potentials* 36.1 (jan. 2017), p. 14-18. issn : 0278-6648. doi : 10.1109/MPOT.2016.2614754. url : http://ieeexplore.ieee.org/document/7814367/.

[3] Saki Asakawa, João Guerreiro, Dragan Ahmetovic, Kris M. Kitani et Chieko Asakawa. “The present and future of museum accessibility for people with visual impairments”. In : *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*. 2018, p. 382-384.

[4] Tayfun Lloyd-Esenkaya, Vanessa Lloyd-Esenkaya, Eamonn O’Neill et Michael J. Proulx. “Multisensory inclusive design with sensory substitution”. en. In : *Cognitive Research: Principles and Implications* 5.1 (août 2020), p. 37. issn : 2365-7464. doi : 10.1186/s41235-020-00240-7. url : https://doi.org/10.1186/s41235-020-00240-7.

[5] Georgina Kleege. “Modes of touch: modelling haptic engagement”. In : *Beyond the Visual: Multisensory modes of beholding art*. Publication Title: UCL Press: London, UK. (2025). London, UK : UCL Press, août 2025, p. 13-19. url : https://doi.org/10.14324/111.9781800088856.

[6] J. Kevin O’Regan et Alva Noë. “A sensorimotor account of vision and visual consciousness”. In : *Behavioral and brain sciences* 24.5 (2001). Publisher: Cambridge University Press, p. 939-973. url : https://www.cambridge.org/core/journals/behavioral-and-brain-sciences/article/sensorimotor-account-of-vision-and-visual-consciousness/BA1638CB7389102A12B336CE687EC270 (visité le 17/09/2025).

[7] Alva Noë. *Action in perception*. MIT press, 2004. url : https://books.google.com/books?hl=fr&lr=&id=kFKvU2hPhxEC&oi=fnd&pg=PP11&dq=action+in+perception&ots=F1xNFjRpr0&sig=uS1XXZqPDNzGPo5E7Z2wkv055UQ (visité le 17/09/2025).

[8] Malika Auvray. “Bergson : une théorie sensorimotrice de la perception”. In : (jan. 2003).

[9] Charles Lenay. “Enaction, externalisme et suppléance perceptive”. fre. In : *Intellectica* 43.1 (2006). Publisher: Persée - Portail des revues scientifiques en SHS, p. 27-52. doi : 10.3406/intel.2006.1326. url : https://www.persee.fr/doc/intel_0769-4113_2006_num_43_1_1326.

[10] Malika Auvray, Sylvain Hanneton et J. Kevin O'Regan. "Learning to perceive with a visuo—auditory substitution system: localisation and object recognition with 'The Voice'". In : *Perception* 36.3 (2007). Publisher: SAGE Publications Sage UK: London, England, p. 416-430.

[11] Paul Bach-y-Rita et Stephen W. Kercel. "Sensory substitution and the human–machine interface". In : *Trends in cognitive sciences* 7.12 (2003). Publisher: Elsevier, p. 541-546. URL : https://www.cell.com/trends/cognitive-sciences/abstract/S1364-6613(03)00290-0 (visité le 17/09/2025).

[12] Marion Chottin et Hannah Thompson. "Blindness Gain as World-Making : Creative Audiodescription as a new "partage du sensible"". In : *L'esprit créateur*. Disability's Worldmaking: Pasts and Futures 61.4 (2021). Publisher: University of Minnesota / Johns Hopkins University Press, p. 32-44. DOI : 10.1353/esp.2021.0045. URL : https://hal.archives-ouvertes.fr/hal-03381142.

[13] Hannah Thompson, Vanessa Warne et Marion Chottin. "Blindness gain and the arts". In : *Beyond the Visual: Multisensory modes of beholding art*. Publication Title: UCL Press: London, UK. (2025). London, UK : UCL Press, août 2025. URL : https://doi.org/10.14324/111.9781800088856.

[14] S. Uboldi, A. Bortolotti, G. Candeloro, A. Marasco, F. Sardella, M. Tartari et P. L. Sacco. "From Touch to Mental Imagery: The Embodied Aesthetic Experience of Late-Blind People Engaged in the Tactile Exploration of Enrico Castellani's Pseudo-Braille Surface". en. In : *Culture, Medicine, and Psychiatry* (mars 2025). ISSN : 0165-005X, 1573-076X. DOI : 10.1007/s11013-025-09904-9. URL : https://link.springer.com/10.1007/s11013-025-09904-9.

[15] Rachel Hutchinson et Alison F. Eardley. "Inclusive museum audio guides: 'guided looking' through audio description enhances memorability of artworks for sighted audiences". In : *Museum Management and Curatorship* 36.4 (juill. 2021). Publisher: Routledge _eprint: https://doi.org/10.1080/09647775.2021.1891563, p. 427-446. ISSN : 0964-7775. DOI : 10.1080/09647775.2021.1891563. URL : https://doi.org/10.1080/09647775.2021.1891563.

[16] Jun Dong Cho. "A Study of Multi-Sensory Experience and Color Recognition in Visual Arts Appreciation of People with Visual Impairment". en. In : *MDPI Electronics* 10.4 (jan. 2021). Number: 4 Publisher: Multidisciplinary Digital Publishing Institute, p. 470. ISSN : 2079-9292. DOI : 10.3390/electronics10040470. URL : https://www.mdpi.com/2079-9292/10/4/470.

[17] Andreas Reichinger, Stefan Maierhofer et Werner Purgathofer. "High-quality tactile paintings". en. In : *Journal on Computing and Cultural Heritage* 4.2 (nov. 2011), p. 1-13. ISSN : 1556-4673, 1556-4711. DOI : 10.1145/2037820.2037822. URL : https://dl.acm.org/doi/10.1145/2037820.2037822.

[18] Katerine Romeo, Marion Chottin, Pierre Ancet et Edwige Pissaloux. "Access to Artworks and Its Mediation by and for Visually Impaired Persons". en. In : *Computers Helping People with Special Needs*. Sous la dir. de Klaus Miesenberger et Georgios

Kouroupetroglou. Lecture Notes in Computer Science. Cham : Springer International Publishing, 2018, p. 233-236. ISBN : 978-3-319-94274-2. DOI : 10.1007/978-3-319-94274-2_32.

[19] Joel Snyder. *The visual made verbal: A comprehensive training manual and guide to the history and applications of audio description*. Æ Academic Publishing, 2020. URL : https://books.google.com/books?hl=fr&lr=&id=WHorEAAAQBAJ&oi=fnd&pg=PR9&dq=The+visual+made+verbal+snyder&ots=kJZOUGpUWZ&sig=OpbOIvZ86EglTkJ1FwCwZawqw1o.

[20] Virginio Cantoni, Luca Lombardi, Alessandra Setti, Stanislav Gyoshev, Dimitar Karastoyanov et Nikolay Stoimenov. "Art masterpieces accessibility for blind and visually impaired people". In : *ICCHP 2018 (Int. Conf. on Computers Helping People with Special Needs)*. Springer, 2018, p. 267-274.

[21] Marion Chottin. *La peinture au prisme de la cécité. Parcours muséal et accessibilité audio-tactile ["Painting through the Prism of Blindness. Museum Tour and Audio-Tactile Accessibility"]*. 2018. URL : http://tristan.u-bourgogne.fr/CGC/manifestations/18_19/18_11_08-09.html.

[22] Francesca Raffi. "Full Access to Cultural Spaces (FACS): Mapping and evaluating museum access services using mobile eye-tracking technology". In : *Ars Aeterna* 9.2 (2017), p. 18-38.

[23] Edwige Pissaloux et Katerine Romeo. *Etat des lieux des médiations techn. dans les musées EU*. French. Workshop. Dijon, 2017.

[24] Marion Chottin. "Proposition d'une charte pour les audiodescriptions de peintures, broderies et tapisserie". In : *2nd Workshop on Digital imaging, heritage and accessibility for all*. Londres, United Kingdom : Hannah Thompson et Katerine Romeo, déc. 2024. URL : https://hal.science/hal-04836416 (visité le 09/07/2025).

[25] Rachel S. Hutchinson et Alison F. Eardley. "Museum audio description: the problem of textual fidelity". In : *Perspectives* 27.1 (2019). Publisher: Taylor & Francis, p. 42-57.

[26] Alison F. Eardley, Louise Fryer, Rachel Hutchinson, Matthew Cock, Peter Ride et Joselia Neves. "Enriched audio description: Working towards an inclusive museum experience". In : *Inclusion, disability and culture*. Springer, 2017, p. 195-207.

[27] Alison F. Eardley, Vanessa E. Jones, Lindsay Bywood, Hannah Thompson et Deborah Husbands. "The W- ICAD model: Redefining museum access through the Workshop for Inclusive Co-created Audio Description". en. In : *Curator: The Museum Journal* 68.1 (jan. 2025), p. 219-242. ISSN : 0011-3069, 2151-6952. DOI : 10.1111/cura.12649. URL : https://onlinelibrary.wiley.com/doi/10.1111/cura.12649 (visité le 18/04/2025).

[28] Josélia Neves. "Multi-sensory approaches to (audio) describing the visual arts". In : *MonTI. Monografías de Traducción e Interpretación* 4 (2012). Publisher: Universitat de València, p. 277-293.

[29] Maria Clara de Almeida, Filipe Herkenhoff Carijó et Virgínia Kastrup. “Por uma estética tátil: sobre a adaptação de obras de artes plásticas para deficientes visuais”. In : *Fractal: Revista de psicologia* 22 (2010). Publisher: SciELO Brasil, traduction in french by D. Valente, p. 85-100.

[30] Andreas Reichinger, Anton L. Fuhrmann, Stefan Maierhofer et Werner Purgathofer. “A Concept for Re-useable Interactive Tactile Reliefs”. In : *null* (2016). doi : 10.1007/978-3-319-41267-2_15.

[31] Luis Cavazos Quero, Jorge Iranzo Bartolomé, Seonggu Lee, En Han, Sunhee Kim et Jundong Cho. “An Interactive Multimodal Guide to Improve Art Accessibility for Blind People”. en. In : *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*. Galway Ireland : ACM, oct. 2018, p. 346-348. isbn : 978-1-4503-5650-3. doi : 10.1145/3234695.3241033. url : https://dl.acm.org/doi/10.1145/3234695.3241033.

[32] Jorge Iranzo Bartolome, Luis Cavazos Quero, Sunhee Kim, Myung-Yong Um et Jundong Cho. “Exploring Art with a Voice Controlled Multimodal Guide for Blind People”. en. In : *Proceedings of the Thirteenth International Conference on Tangible, Embedded, and Embodied Interaction*. Tempe Arizona USA : ACM, mars 2019, p. 383-390. isbn : 978-1-4503-6196-5. doi : 10.1145/3294109.3300994. url : https://dl.acm.org/doi/10.1145/3294109.3300994.

[33] Meredith Ringel Morris, Jazette Johnson, Cynthia L. Bennett et Edward Cutrell. “Rich representations of visual content for screen reader users”. In : *Proc. of the 2018 CHI (conf. on Human Factors in Computing Systems)*. ACM, 2018, p. 1-11.

[34] Nahyun Kwon, Youngji Koh et Uran Oh. “Supporting object-level exploration of artworks by touch for people with visual impairments”. In : *The 21st International ACM SIGACCESS Conference on Computers and Accessibility*. 2019, p. 600-602.

[35] Nahyun Kwon, Yunjung Lee et Uran Oh. “Supporting a crowd-powered accessible online art gallery for people with visual impairments: a feasibility study”. en. In : *Universal Access in the Information Society* (mai 2021). issn : 1615-5297. doi : 10.1007/s10209-021-00814-2. url : https://doi.org/10.1007/s10209-021-00814-2.

[36] M. S. Jabbar, J. Shin et J. D. Cho. *AI Ekphrasis: Multi-Modal Learning with Foundation Models for Fine-Grained Poetry Retrieval. Electronics 2022, 11, 1275*. 2022. url : https://www.academia.edu/download/107233624/pdf.pdf.

[37] Ohini Kafui Toffa et M. Mignotte. “Dataset and semantic based-approach for image sonification”. In : *Multimedia Tools and Applications* (2022). doi : 10.1007/s11042-022-12914-z.

[38] Jaewook Lee, Jaylin Herskovitz, Yi-Hao Peng et Anhong Guo. “ImageExplorer: Multi-Layered Touch Exploration to Encourage Skepticism Towards Imperfect AI-Generated Image Captions”. In : *CHI Conference on Human Factors in Computing Systems*. 2022, p. 1-15.

[39] Jun-Dong Cho et Yong Lee. “ColorPoetry: Multi-Sensory Experience of Color with Poetry in Visual Arts Appreciation of Persons with Visual Impairment”. en. In : *MDPI Electronics* 10.9 (jan. 2021). Number: 9 Publisher: Multidisciplinary Digital Publishing Institute, p. 1064. ISSN : 2079-9292. DOI : 10.3390/electronics10091064. URL : https://www.mdpi.com/2079-9292/10/9/1064.

[40] Jun Dong Cho, Jun-Dong Cho, Jaeho Jeong, Ji Hye Kim et Hoonsuk Lee. “Sound Coding Color to Improve Artwork Appreciation by People with Visual Impairments”. In : *MDPI Electronics* (2020). DOI : 10.3390/electronics9111981.

[41] Yong Lee, Chung-Heon Lee et Jun Dong Cho. “3D Sound Coding Color for the Visually Impaired”. en. In : *MDPI Electronics* 10.9 (jan. 2021). Number: 9 Publisher: Multidisciplinary Digital Publishing Institute, p. 1037. ISSN : 2079-9292. DOI : 10.3390/electronics10091037. URL : https://www.mdpi.com/2079-9292/10/9/1037.

[42] Dragan Ahmetovic, Nahyun Kwon, Uran Oh, Cristian Bernareggi et Sergio Mascetti. “Touch Screen Exploration of Visual Artwork for Blind People”. en. In : *Proceedings of the Web Conference 2021*. Ljubljana Slovenia : ACM, avr. 2021, p. 2781-2791. ISBN : 978-1-4503-8312-7. DOI : 10.1145/3442381.3449871. URL : https://dl.acm.org/doi/10.1145/3442381.3449871.

[43] Tyler Ferro. “Automatic Image Processing and Conversion to Tactile Graphics”. PhD Thesis. Virginia Commonwealth University, jan. 2018. URL : https://scholarscompass.vcu.edu/etd/5469.

[44] Chandrika Jayant, Matt Renzelmann, Dana Wen, Satria Krisnandi, Richard Ladner et Dan Comden. “Automated tactile graphics translation: in the field”. In : *Proc. of the 9th Int. ACM SIGACCESS Conf. on Computers and Accessibility*. 2007, p. 75-82.

[45] Abigale Stangl, Chia-Lo Hsu et Tom Yeh. “Transcribing across the senses: community efforts to create 3D printable accessible tactile pictures for young children with visual impairments”. In : *Proc. of the 17th Int. ACM SIGACCESS Conf. on Computers & Accessibility*. 2015, p. 127-137.

[46] Yary Volpe, Rocco Furferi, Lapo Governi et Giovanna Tennirelli. “Computer-based methodologies for semi-automatic 3D model generation from paintings”. In : *International Journal of Computer Aided Engineering and Technology* 6.1 (2014). Publisher: Inderscience Publishers Ltd, p. 88-112.

[47] Yukino Kobayashi, Manabi Miyagi et Eri Ishikawa. “Introducing Art to People with Visual Impairment through Tactile Pictures: A Survey of Art Museums in Japan”. en. In : *ICCHP-AAATE 2022 Open Access Compendium "Assistive Technology* Accessibility and (2022). Artwork Size: 7 pages Medium: PDF Publisher: Association ICCHP, 7 pages. DOI : 10.35011/ICCHP-AAATE22-P1-05. URL : https://epub.jku.at/obvulioa/content/titleinfo/7945366.

[48] Jaeeun Shin, Jun-Dong Cho, Sangwon Lee, Sangwon Lee et Sangwon Lee. “Please Touch Color: Tactile-Color Texture Design for The Visually Impaired”. In : *CHI Extended Abstracts* (2020). DOI : 10.1145/3334480.3383003.

[49] Jun Dong Cho, Luis Cavazos Quero, Jorge Iranzo Bartolome, Do Won Lee, Uran Oh et Inae Lee. “Tactile colour pictogram to improve artwork appreciation of people with visual impairments”. In : *Color Research & Application* 46.1 (2021). Publisher: Wiley Online Library, p. 103-116.

[50] Jaeeun Shin, Jundong Cho et Sangwon Lee. “Tactile-Color System for Accessibility of Color Education: 2.5D UV Printed Supplementary Material for Visually Impaired Students”. In : *2021 15th International Conference on Ubiquitous Information Management and Communication (IMCOM)*. Jan. 2021, p. 1-7. doi : 10.1109/IMCOM51814.2021.9377433.

[51] Charles Lenay, Olivier Gapenne, Sylvain Hanneton, Catherine Marque et Christelle Genouëlle. “Chapter 16. Sensory substitution: Limits and perspectives”. en. In : *Advances in Consciousness Research*. Sous la dir. d’Yvette Hatwell, Arlette Streri et Edouard Gentaz. T. 53. Amsterdam : John Benjamins Publishing Company, 2003, p. 275-292. isbn : 978-90-272-5185-5. doi : 10.1075/aicr.53.22len. url : https://benjamins.com/catalog/aicr.53.22len.

[52] Shachar Maidenbaum, Sami Abboud et Amir Amedi. “Sensory substitution: Closing the gap between basic research and widespread practical visual rehabilitation”. In : *Neuroscience & Biobehavioral Reviews* 41 (2014). Publisher: Elsevier, p. 3-15.

[53] Carter C. Collins, Frank A. Saunders, Benjamin White et Lawrence Scadden. “Vision substitution by tactile image projection”. In : *Nature* 221.5184 (1969). Publisher: Nature Publishing Group, p. 963-964.

[54] Malika Auvray et Erik Myin. “Perception With Compensatory Devices: From Sensory Substitution to Sensorimotor Extension”. en. In : *Cognitive Science* 33.6 (2009). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1551-6709.2009.01040.x, p. 1036-1058. issn : 1551-6709. doi : 10.1111/j.1551-6709.2009.01040.x. url : https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1551-6709.2009.01040.x.

[55] Karon E MacLean. “Designing affective haptic experience for wellness and social communication: where designers need affective neuroscience and psychology”. In : *Current Opinion in Behavioral Sciences* 45 (juin 2022), p. 101113. issn : 2352-1546. doi : 10.1016/j.cobeha.2022.101113. url : https://www.sciencedirect.com/science/article/pii/S2352154622000195.

[56] Karon E. MacLean, Oliver S. Schneider et Hasti Seifi. “Multisensory haptic interactions: understanding the sense and designing for it”. In : *The Handbook of Multimodal-Multisensor Interfaces: Foundations, User Modeling, and Common Modality Combinations - Volume 1*. T. 14. Association for Computing Machinery et Morgan & Claypool, 2017, p. 97-142. isbn : 978-1-970001-67-9. url : https://doi.org/10.1145/3015783.3015788.

[57] Louis H. Goldish et Harry E. Taylor. “The Optacon: A Valuable Device for Blind Persons”. en. In : *Journal of Visual Impairment & Blindness* 68.2 (fév. 1974), p. 49-56. issn : 0145-482X, 1559-1476. doi : 10.1177/0145482X7406800201. url : https://journals.sagepub.com/doi/10.1177/0145482X7406800201.

[58] Mounia Ziat, Olivier Gapenne, Charles Lenay et John Stewart. “Zooming experience in the haptic modality”. In : *4th International Conference on Enactive Interfaces 2007*. 2007, p. 305. URL : https://www.academia.edu/download/50471698/Topophonies_and_score_navigation_in_ENIG20161121-20758-i6te3g.pdf#page=305.

[59] Seondae Kim, Eun Soo Park, Eun-Soo Park et Eun-Seok Ryu. “Multimedia Vision for the Visually Impaired through 2D Multiarray Braille Display”. In : *MDPI Applied Sciences* (2019). DOI : 10.3390/app9050878.

[60] Sile O’Modhrain, Nicholas A. Giudice, John A. Gardner et Gordon E. Legge. “Designing Media for Visually-Impaired Users of Refreshable Touch Displays: Possibilities and Pitfalls”. In : *IEEE Transactions on Haptics* 8.3 (juill. 2015). Conference Name: IEEE Transactions on Haptics, p. 248-257. ISSN : 2329-4051. DOI : 10.1109/TOH.2015.2466231.

[61] Grégory Petit, Aude Dufresne, Vincent Lévesque et Vincent Hayward. “Exploration multimodale d’images pour des utilisateurs ayant une déficience visuelle”. fr. In : *Sciences et Technologies pour le Handicap* 2.2 (déc. 2008), p. 175-186. ISSN : 19602081. DOI : 10.3166/sth.2.175-186. URL : http://sth.revuesonline.com/article.jsp?articleId=13289.

[62] Victoria E. Hribar et Dianne TV Pawluk. “A tactile-thermal display for haptic exploration of virtual paintings”. In : *The Proc. of the 13th int. ACM SIGACCESS conf. on Computers and accessibility*. 2011, p. 221-222.

[63] Jorge Iranzo Bartolomé, Luis Cavazos Quero, Jun-Dong Cho et Sunggi Jo. “Exploring Thermal Interaction for Visual Art Color Appreciation for the Visually Impaired People”. In : *2020 International Conference on Electronics, Information, and Communication (ICEIC)* (2020). DOI : 10.1109/iceic49074.2020.9051245.

[64] Jorge Iranzo Bartolome, Gilsang Cho et Jun-Dong Cho. “Multi-Sensory Color Expression with Sound and Temperature in Visual Arts Appreciation for People with Visual Impairment”. en. In : *MDPI Electronics* 10.11 (jan. 2021). Number: 11 Publisher: Multidisciplinary Digital Publishing Institute, p. 1336. ISSN : 2079-9292. DOI : 10.3390/electronics10111336. URL : https://www.mdpi.com/2079-9292/10/11/1336.

[65] Luis Cavazos Quero, Chung-Heon Lee et Jun-Dong Cho. “Multi-sensory color code based on sound and scent for visual art appreciation”. In : *MDPI Electronics* 10.14 (2021). Publisher: MDPI, p. 1696.

[66] Cheng Xu, Ali Israr, Ivan Poupyrev, Olivier Bau et Chris Harrison. “Tactile display for the visually impaired using TeslaTouch”. In : *CHI ’11 Extended Abstracts on Human Factors in Computing Systems*. CHI EA ’11. New York, NY, USA : Association for Computing Machinery, 2011, p. 317-322. ISBN : 978-1-4503-0268-5. DOI : 10.1145/1979742.1979705. URL : https://doi.org/10.1145/1979742.1979705.

[67] Michel Amberg, Frédéric Giraud, Betty Semail, Paolo Olivo, Géry Casiez et Nicolas Roussel. “STIMTAC: a tactile input device with programmable friction”. In : *Proceedings of the 24th annual ACM symposium adjunct on User interface software and technology*. 2011, p. 7-8.

[68] Yalan Luo, Weiyue Lin, Yuhan Liu, Xiaomei Nie, Xiang Qian et Hanyu Guo. “We-see: Digital Cultural Heritage Interpretation for Blind and Low Vision People”. en. In : *Human-Computer Interaction – INTERACT 2023*. Sous la dir. de José Abdelnour Nocera, Marta Kristín Lárusdóttir, Helen Petrie, Antonio Piccinno et Marco Winckler. Cham : Springer Nature Switzerland, 2023, p. 123-131. isbn : 978-3-031-42280-5. doi : 10.1007/978-3-031-42280-5_8.

[69] Grégory Petit, Aude Dufresne, Vincent Levesque, Vincent Hayward et Nicole Trudeau. “Graphisme tactile appliqué aux illustrations de manuels scolaires à l’usage d’enfants ayant une déficience visuelle”. en. In : *Proceedings of the 20th Conference on l’Interaction Homme-Machine*. Metz France : ACM, sept. 2008, p. 73-80. isbn : 978-1-60558-285-6. doi : 10.1145/1512714.1512729. url : https://dl.acm.org/doi/10.1145/1512714.1512729.

[70] Grégory Petit, Aude Dufresne, Vincent Levesque, Vincent Hayward et Nicole Trudeau. “Refreshable tactile graphics applied to schoolbook illustrations for students with visual impairment”. en. In : *Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility*. Halifax Nova Scotia Canada : ACM, oct. 2008, p. 89-96. isbn : 978-1-59593-976-0. doi : 10.1145/1414471.1414489. url : https://dl.acm.org/doi/10.1145/1414471.1414489.

[71] G. Campion, Qi Wang et V. Hayward. “The Pantograph Mk-II: a haptic instrument”. In : *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*. ISSN: 2153-0866. Août 2005, p. 193-198. doi : 10.1109/IROS.2005.1545066.

[72] Qi Wang et V. Hayward. “Compact, Portable, Modular, High-performance, Distributed Tactile Transducer Device Based on Lateral Skin Deformation”. en. In : *2006 14th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*. Alexandria, VA, USA : IEEE, 2006, p. 67-72. isbn : 978-1-4244-0226-7. doi : 10.1109/HAPTIC.2006.1627091. url : http://ieeexplore.ieee.org/document/1627091/.

[73] Radu Comes. “Haptic devices and tactile experiences in museum exhibitions”. In : *Journal of Ancient History and Archaeology* 3.4 (2016). url : http://jaha.org.ro/index.php/JAHA/article/view/205 (visité le 03/10/2025).

[74] Margaret L. McLaughlin, Gaurav Sukhatme, Cyrus Shahabi, Joao Hespanha, Antonio Ortega et Gerard Medioni. “The haptic museum”. In : *Proceedings of the EVA 2000 conference on electronic imaging and the visual arts*. 2000, p. 1-6. url : https://www.academia.edu/download/87357196/eva2000.pdf (visité le 03/10/2025).

[75] Roberto Vaz, Diamantino Freitas et António Coelho. “Blind and Visually Impaired Visitors’ Experiences in Museums: Increasing Accessibility through Assistive Technologies.” In : *International Journal of the Inclusive Museum* 13.2 (2020).

[76] Andreas Reichinger, Svenja Schröder, Christian Löw, Sam Sportun, Peter Reichl et Werner Purgathofer. “Spaghetti, Sink and Sarcophagus: Design Explorations of Tactile Artworks for Visually Impaired People”. In : *NordiCHI* (2016). doi : 10.1145/2971485.2996471.

[77] Stanislav Gyoshev, Dimitar Karastoyanov, Nikolay Stoimenov, Virginio Cantoni, Luca Lombardi et Alessandra Setti. “Exploiting a graphical Braille display for art masterpieces”. In : *Int. Conf. on Computers Helping People with Special Needs*. Springer, 2018, p. 237-245.

[78] OD Sasanka Fonseka, N. Wedasinghe et Nirosha Wedasinghe. “-Blind Draw-A software solution for image identification and artistic skills for visually impaired people using braille”. In : *2018 3rd International Conference on Information Technology Research (ICITR)* (2018). doi : 10.1109/icitr.2018.8736137.

[79] Ravi Rastogi, TV Dianne Pawluk et Jessica Ketchum. “Intuitive tactile zooming for graphics accessed by individuals who are blind and visually impaired”. In : *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 21.4 (2013). Publisher: IEEE, p. 655-663.

[80] Calle Sjostrom. “Designing haptic computer interfaces for blind people”. In : *Proceedings of the Sixth International Symposium on Signal Processing and its Applications (Cat. No. 01EX467)*. T. 1. IEEE, 2001, p. 68-71. url : https://ieeexplore.ieee.org/abstract/document/949777/.

[81] Francesco Buonamici, Monica Carfagni, Rocco Furferi, Lapo Governi et Yary Volpe. “Are we ready to build a system for assisting blind people in tactile exploration of bas-reliefs?” In : *Sensors* 16.9 (2016). Publisher: Multidisciplinary Digital Publishing Institute, p. 1361.

[82] Andreas Reichinger et Joanna Wood. “Pictures in Your Mind: Using Interactive Gesture-Controlled Reliefs to Explore Art”. In : *ACM Transactions on Accessible Computing (TACCESS)* (2018). doi : https://doi.org/10.1145/3155286.

[83] Luis Cavazos Quero, Jorge Iranzo Bartolomé et Jundong Cho. “Accessible Visual Artworks for Blind and Visually Impaired People: Comparing a Multimodal Approach with Tactile Graphics”. en. In : *Electronics* 10.3 (jan. 2021). Number: 3 Publisher: Multidisciplinary Digital Publishing Institute, p. 297. issn : 2079-9292. doi : 10.3390/electronics10030297. url : https://www.mdpi.com/2079-9292/10/3/297.

[84] Yassine Fadlaoui et Dominique Archambault. “Haptic and interactive illustrations for visually impaired children: A literature review”. EN. In : *Technology and Disability* 37.3 (août 2025). Publisher: SAGE Publications, p. 229-236. issn : 1055-4181. doi : 10.1177/10554181251355430. url : https://doi.org/10.1177/10554181251355430.

[85] Amal Ali Ammar, Olivier Gapenne, Elodie Blomme et Katia Rovira. “Analyse de l’exploration tactile sur support traditionnel chez la personne aveugle et conception de l’interface de lecture Tactos.” In : *Rev. d’Intelligence Artif.* 19.1-2 (2005), p. 339-354. URL : https://www.academia.edu/download/69510303/Analyse_de_lexploration_tactile_sur_sup20210912-14653-1txk619.pdf.

[86] Susan J. Lederman et Roberta L. Klatzky. “Extracting object properties through haptic exploration”. In : *Acta psychologica* 84.1 (1993). Publisher: Elsevier, p. 29-40.

[87] Soledad Ballesteros, Dionisio Manga et Jose Manuel Reales. “Haptic discrimination of bilateral symmetry in 2-dimensional and 3-dimensional unfamiliar displays”. en. In : *Perception & Psychophysics* 59.1 (jan. 1997). Publisher: Springer Science and Business Media LLC, p. 37-50. ISSN : 0031-5117, 1532-5962. DOI : 10.3758/bf03206846. URL : http://link.springer.com/10.3758/BF03206846.

[88] Maarten W A Wijntjes, Thijs van Lienen, Ilse M Verstijnen et Astrid M L Kappers. “The Influence of Picture Size on Recognition and Exploratory Behaviour in Raised-Line Drawings”. EN. In : *Perception* 37.4 (avr. 2008). Publisher: SAGE Publications Ltd STM, p. 602-614. ISSN : 0301-0066. DOI : 10.1068/p5714. URL : https://doi.org/10.1068/p5714.

[89] Oriana Orlandi. “La compréhension des images tactiles chez les enfants porteurs d’un handicap visuel”. These de doctorat. Dijon, juill. 2015. URL : https://theses.fr/2015DIJOL006.

[90] Yvette Hatwell, Christian Osiek et Véronique Jeanneret. “L’exploration perceptive tactile d’un ensemble d’objets chez l’enfant et chez l’adulte”. fre. In : *L’Année psychologique* 73.2 (1973). Company: Persée - Portail des revues scientifiques en SHS Distributor: Persée - Portail des revues scientifiques en SHS Institution: Persée - Portail des revues scientifiques en SHS Label: Persée - Portail des revues scientifiques en SHS Publisher: Université René Descartes, Paris 5, p. 419-441. DOI : 10.3406/psy.1973.27996. URL : https://www.persee.fr/doc/psy_0003-5033_1973_num_73_2_27996.

[91] Dannyelle Valente, Amaya Palama, Jennifer Malsert, Guillemette Bolens et Edouard Gentaz. “Adults’ visual recognition of actions simulations by finger gestures (ASFGs) produced by sighted and blind individuals”. In : *Plos one* 14.3 (2019). Publisher: Public Library of Science San Francisco, CA USA, e0214371. URL : https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0214371.

[92] Dannyelle Valente, Amaya Palama et Edouard Gentaz. “Exploring 3D miniatures with action simulations by finger gestures: Study of a new embodied design for blind and sighted children”. In : *PloS one* 16.2 (2021). Publisher: Public Library of Science San Francisco, CA USA, e0245472. URL : https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0245472.

[93] Dannyelle Valente, Lola Chennaz, Dominique Archambault, Solène Négrerie, Sophie Blain, Anna Rita Galiano et Edouard Gentaz. “Comprehension of a multimodal book by children with visual impairments”. en. In : *British Journal of Visual Impairment* 42.1 (jan. 2024), p. 276-286. ISSN : 0264-6196, 1744-5809. DOI :

10.1177/02646196231172071. URL : https://journals.sagepub.com/doi/10.1177/02646196231172071.

[94] Loïc Deschamps, Charles Lenay, Katia Rovira, Université de Technologie de Compiègne et École doctorale Sciences pour l'ingénieur. *Suppléance perceptive et cognition sociale*. français. 2013. URL : https://bibliotheque.utc.fr/Default/doc/SYRACUSE/336949/suppleance-perceptive-et-cognition-sociale.

[95] Mounia Ziat, Charles Lenay, Olivier Gapenne, John Stewart, Amal Ali Ammar et Dominique Aubert. "Perceptive Supplementation for an Access to Graphical Interfaces". en. In : *Universal Acess in Human Computer Interaction. Coping with Diversity*. Sous la dir. de David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu Sudan, Demetri Terzopoulos, Doug Tygar, Moshe Y. Vardi, Gerhard Weikum et Constantine Stephanidis. T. 4554. Series Title: Lecture Notes in Computer Science. Berlin, Heidelberg : Springer Berlin Heidelberg, 2007, p. 841-850. ISBN : 978-3-540-73279-2_94. DOI : 10.1007/978-3-540-73279-2_94. URL : http://link.springer.com/10.1007/978-3-540-73279-2_94.

[96] Katia Rovira, Olivier Gapenne et Aurélie Vallée. "Suppléance perceptive chez l'adolescent aveugle : stratégies individuelles, perception et catégorisation de forme". fr. In : *Enfance* 1.1 (2014). Publisher: NecPlus Section: Sciences de l'éducation, p. 89-106. ISSN : 0013-7545. DOI : 10.3917/enf1.141.0089. URL : https://shs.cairn.info/revue-enfance2-2014-1-page-89.

[97] Emilie Rumilly. "Stratégies du regard de l'enfant et photo-oculographie binoculaire: faisabilité et pertinence". fr. Pages: non renseigné. other. UHP - Université Henri Poincaré, oct. 2007. URL : https://hal.univ-lorraine.fr/hal-01734149.

[98] Waseem Safi, Fabrice Maurel, Jean-Marc Routoure, Pierre Beust, Michèle Molina, Coralie Sann et Jessica Guilbert. "Blind Navigation of Web Pages through Vibrotactile Feedbacks". en. In : *25th ACM Symposium on Virtual Reality Software and Technology*. Parramatta NSW Australia : ACM, nov. 2019, p. 1-1. ISBN : 978-1-4503-7001-1. DOI : 10.1145/3359996.3364758. URL : https://dl.acm.org/doi/10.1145/3359996.3364758.

[99] Fabrice Maurel et Waseem Safi. "La TactiNET: toucher le Web... Pour mieux l'entendre". In : *27ème conférence francophone sur l'Interaction Homme-Machine*. ACM, 2015, w2. URL : https://hal.science/hal-01219067/.

[100] Francis Edeline, Jean-Marie Klinkenberg, Philippe Minguet et Groupe Mu. "Traité du signe visuel: pour une rhétorique de l'image". In : (1992). URL : https://philpapers.org/rec/MUTDS.

[101] David H. Hubel et Torsten N. Wiesel. "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex". In : *The Journal of physiology* 160.1 (1962), p. 106. URL : https://pmc.ncbi.nlm.nih.gov/articles/PMC1359523/.

[102] Bruno A. Olshausen et David J. Field. “Emergence of simple-cell receptive field properties by learning a sparse code for natural images”. In : *Nature* 381.6583 (1996). Publisher: Nature Publishing Group UK London, p. 607-609. URL : https://www.nature.com/articles/381607a0.

[103] Volker Koch, Angelika J. Lückert, Thorsten Schwarz, Petra von Both et Peter Diziol. “Haptic Paintings: Using Rapid Prototyping Technologies to Grant Visually Impaired Persons Access to Paintings, Sculptures, Graphics and Architecture”. In : (2012). Publisher: CUMINCAD.

[104] Andreas Reichinger, Anton Fuhrmann, Stefan Maierhofer et Werner Purgathofer. “Gesture-Based Interactive Audio Guide on Tactile Reliefs”. en. In : *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility*. Reno Nevada USA : ACM, oct. 2016, p. 91-100. ISBN : 978-1-4503-4124-0. DOI : 10.1145/2982142.2982176. URL : https://dl.acm.org/doi/10.1145/2982142.2982176.

[105] Luis Cavazos Quero, Jorge Iranzo Bartolomé et Jun-Dong Cho. “Accessible Visual Artworks for Blind and Visually Impaired People: Comparing a Multimodal Approach with Tactile Graphics”. In : *MDPI Electronics* (2021). DOI : 10.3390/electronics10030297.

[106] Jorge Iranzo Bartolomé, Jun-Dong Cho, Luis Cavazos Quero, Sunggi Jo et Gilsang Cho. “Thermal Interaction for Improving Tactile Artwork Depth and Color-Depth Appreciation for Visually Impaired People”. In : *MDPI Electronics* (2020). DOI : 10.3390/electronics9111939.

[107] Marjorie Redon, Matthieu Pizenberg, Yvain Quéau et Abderrahim Elmoataz. “3d surface approximation of the entire bayeux tapestry for improved pedagogical access”. In : *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2023, p. 1593-1602. URL : https://openaccess.thecvf.com/content/ICCV2023W/e-Heritage/html/Redon_3D_Surface_Approximation_of_the_Entire_Bayeux_Tapestry_for_Improved_ICCVW_2023_paper.html.

[108] Marjorie Redon, Matthieu Pizenberg, Yvain Quéau et Abderrahim Elmoataz. “Vers un accès plus inclusif à la Tapisserie de Bayeux par la génération semi-automatique d’objets 3D”. In : *ORASIS, 19ièmes journées francophones des jeunes chercheurs en vision par ordinateur*. Carqueiranne, France, mai 2023. URL : https://hal.science/hal-04075676.

[109] Matthieu Pizenberg, Abderrahim Elmoataz et Yvain Quéau. “Image Processing for Cultural Heritage Accessibility: Digitizing the Bayeux Tapestry”. In : (2022).

[110] Youngjun Kim, Hayoung Jeong, Jun-Dong Cho et Jitae Shin. “Construction of a soundscape-based media art exhibition to improve user appreciation experience by using deep neural networks”. In : *MDPI Electronics* 10.10 (2021). Publisher: MDPI, p. 1170.

[111] Francesco de Gioia, Massimiliano Donati, Luca Fanucci, Luca Fanucci et Luca Fanucci. “Automatic Generation of 3D Printable Tactile Paintings for the Visually Impaired”. In : *ApplePies* (2021). DOI : 10.1007/978-3-030-66729-0_10.

[112] Dragan Ahmetovic, Cristian Bernareggi, Kristian Keller et Sergio Mascetti. “MusA: artwork accessibility through augmented reality for people with low vision”. en. In : *Proceedings of the 18th International Web for All Conference*. Ljubljana Slovenia : ACM, avr. 2021, p. 1-9. isbn : 978-1-4503-8212-0. doi : 10.1145/3430263.3452441. url : https://dl.acm.org/doi/10.1145/3430263.3452441.

[113] Kyle Reinholt, Darren Guinness et Shaun K. Kane. “Eyedescribe: Combining eye gaze and speech to automatically create accessible touch screen artwork”. In : *Proceedings of the 2019 ACM International Conference on Interactive Surfaces and Spaces*. 2019, p. 101-112.

[114] Fiona Candlin. “Don’t touch! Hands off! Art, blindness and the conservation of expertise”. In : *Body & Society* 10.1 (2004). Publisher: Sage Publications, p. 71-90.

[115] Leona Holloway, Kim Marriott, Matthew Butler et Alan Borning. “Making Sense of Art: Access for Gallery Visitors with Vision Impairments”. en. In : *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (mai 2019). Conference Name: CHI ’19: CHI Conference on Human Factors in Computing Systems ISBN: 9781450359702 Place: Glasgow Scotland Uk Publisher: ACM, p. 1-12. doi : 10.1145/3290605.3300250. url : https://dl.acm.org/doi/10.1145/3290605.3300250.

[116] Sandra Bardot. “Conception et évaluation de techniques d’interaction non-visuelle basées sur un dispositif personnel : application à l’exploration de cartes numériques par des personnes avec déficience visuelle”. fr. Thèse de doct. Université Paul Sabatier - Toulouse III, juin 2019. url : https://theses.hal.science/tel-02896422.

[117] Tjaša Krivec, Tadeja Muck, Rolanda Fugger Germadnik, Igor Majnarić et Gorazd Golob. “Adapting Artworks for People who Are Blind or Visually Impaired Using Raised Printing”. en. In : *Journal of Visual Impairment & Blindness* 108.1 (jan. 2014). Publisher: SAGE Publications Inc, p. 68-76. issn : 0145-482X. doi : 10.1177/0145482X1410800108. url : https://doi.org/10.1177/0145482X1410800108.

[118] Pingjun Xia. “New advances for haptic rendering: state of the art”. en. In : *The Visual Computer* 34.2 (fév. 2018), p. 271-287. issn : 0178-2789, 1432-2315. doi : 10.1007/s00371-016-1324-y. url : http://link.springer.com/10.1007/s00371-016-1324-y.

[119] Carsten Rother, Vladimir Kolmogorov et Andrew Blake. “"GrabCut": interactive foreground extraction using iterated graph cuts”. In : *ACM SIGGRAPH 2004 Papers*. SIGGRAPH ’04. New York, NY, USA : Association for Computing Machinery, 2004, p. 309-314. isbn : 978-1-4503-7823-9. doi : 10.1145/1186562.1015720. url : https://doi.org/10.1145/1186562.1015720.

[120] William A. Barrett et Eric N. Mortensen. “Interactive live-wire boundary extraction”. en. In : *Medical Image Analysis* 1.4 (sept. 1997), p. 331-341. issn : 13618415. doi : 10.1016/S1361-8415(97)85005-0. url : https://linkinghub.elsevier.com/retrieve/pii/S1361841597850050.

[121] Rocco Furferi, Lapo Governi, Yary Volpe, Luca Puggelli, Niccolò Vanni et Monica Carfagni. “From 2D to 2.5D i.e. from painting to tactile model”. en. In : *Graphical Models* 76.6 (nov. 2014), p. 706-723. issn : 1524-0703. doi : 10.1016/j.gmod.2014.10.001. url : https://www.sciencedirect.com/science/article/pii/S1524070314000526.

[122] Jordan Boaz Rodrigues, Alinne Victoria Martins Ferreira, Ivana Marcia Oliveira Maia, Geraldo Braz Junior, João Dallyson Sousa de Almeida et Anselmo Cardoso de Paiva. “Image Processing of Artworks for Construction of 3D Models Accessible to the Visually Impaired”. In : *Advances in Manufacturing, Production Management and Process Control* (2018). doi : 10.1007/978-3-319-94196-7_23.

[123] Jialu Li, Aiguo Song et Xiaorui Zhang. “Haptic Texture Rendering Using Single Texture Image”. In : *2010 International Symposium on Computational Intelligence and Design*. T. 1. Oct. 2010, p. 7-10. doi : 10.1109/ISCID.2010.9. url : https://ieeexplore.ieee.org/abstract/document/5692650?casa_token=lI8UjQTEHngAAAAA:lVFWupKxlhroww0ucMxKLJOmLb_2DUh-xaVK8nvav-x9nfGdNe7OCkyVUCeqmVXVMG9AvzEcXTY.

[124] Xingzi Zhang et Alexei Sourin. “Image-inspired haptic interaction”. en. In : *Computer Animation and Virtual Worlds* 26.3-4 (2015), p. 311-319. issn : 1546-427X. doi : 10.1002/cav.1643. url : https://onlinelibrary.wiley.com/doi/abs/10.1002/cav.1643.

[125] Shahzad Rasool et Alexei Sourin. “Image-Driven Haptic Rendering”. en. In : *Transactions on Computational Science XXIII: Special Issue on Cyberworlds*. Sous la dir. de Marina L. Gavrilova, C. J. Kenneth Tan, Xiaoyang Mao et Lichan Hong. Lecture Notes in Computer Science. Berlin, Heidelberg : Springer, 2014, p. 58-77. isbn : 978-3-662-43790-2. doi : 10.1007/978-3-662-43790-2_4. url : https://doi.org/10.1007/978-3-662-43790-2_4.

[126] Thomas P. Way et Kenneth E. Barner. “Automatic visual to tactile translation. i. human factors, access methods and image manipulation”. In : *IEEE Transactions on rehabilitation engineering* 5.1 (1997). Publisher: IEEE, p. 81-94.

[127] Sergio E. Hernandez et Kenneth E. Barner. “Tactile imaging using watershed-based image segmentation”. In : *Proceedings of the fourth international ACM conference on Assistive technologies*. Assets ’00. New York, NY, USA : ACM, nov. 2000, p. 26-33. isbn : 978-1-58113-313-4. doi : 10.1145/354324.354332. url : https://doi.org/10.1145/354324.354332.

[128] Saining Xie et Zhuowen Tu. “Holistically-Nested Edge Detection”. In : 2015, p. 1395-1403. url : https://openaccess.thecvf.com/content_iccv_2015/html/Xie_Holistically-Nested_Edge_Detection_ICCV_2015_paper.html.

[129] Ali Souradi, Christele Lecomte, Katerine Romeo, Simon Gay, Marc-Aurele Riviere, Abderrahim El Moataz et Edwige Pissaloux. “Towards the Tactile Discovery of Cultural Heritage with Multi-approach Segmentation”. en. In : *Image and Signal Processing*. Sous la dir. d’Abderrahim El Moataz, Driss Mammass, Alamin Mansouri

et Fathallah Nouboud. Cham : Springer International Publishing, 2020, p. 14-23. isbn : 978-3-030-51935-3. doi : 10.1007/978-3-030-51935-3_2.

[130] Hari Vasudevan et .M Manivannan. “Tangible Images: Runtime Generation of Haptic Textures From Images”. In : *2008 Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*. ISSN: 2324-7355. Mars 2008, p. 357-360. doi : 10.1109/HAPTICS.2008.4479971. url : https://ieeexplore.ieee.org/document/4479971.

[131] Jialu Li, Aiguo Song et Xiaorui Zhang. “Image-based haptic texture rendering”. In : *Proceedings of the 9th ACM SIGGRAPH Conference on Virtual-Reality Continuum and its Applications in Industry*. VRCAI ’10. New York, NY, USA : Association for Computing Machinery, 2010, p. 237-242. isbn : 978-1-4503-0459-7. doi : 10.1145/1900179.1900230. url : https://doi.org/10.1145/1900179.1900230.

[132] Seung-Chan Kim, Ki-Uk Kyung et Dong-Soo Kwon. “Haptic annotation for an interactive image”. In : *Proceedings of the 5th International Conference on Ubiquitous Information Management and Communication*. ICUIMC ’11. New York, NY, USA : Association for Computing Machinery, 2011, p. 1-10. isbn : 978-1-4503-0571-6. doi : 10.1145/1968613.1968675. url : https://doi.org/10.1145/1968613.1968675.

[133] Seung-Chan Kim, Byung-Kil Han, Jeong-Yean Yang et Dong-Soo Kwon. “Interaction with objects inside a media space”. In : *ACM SIGGRAPH ASIA 2010 Posters*. SA ’10. New York, NY, USA : Association for Computing Machinery, 2010, p. 1. isbn : 978-1-4503-0524-2. doi : 10.1145/1900354.1900397. url : https://doi.org/10.1145/1900354.1900397.

[134] Pablo Arbelaez, Michael Maire, Charless Fowlkes et Jitendra Malik. “Contour detection and hierarchical image segmentation”. In : *IEEE transactions on pattern analysis and machine intelligence* 33.5 (2010). Publisher: IEEE, p. 898-916.

[135] Pablo Arbelaez, Michael Maire, Charless Fowlkes et Jitendra Malik. “From contours to regions: An empirical evaluation”. In : *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2009, p. 2294-2301.

[136] Akmalbek Abdusalomov, Mukhriddin Mukhiddinov, Oybek Djuraev, Utkir Khamdamov et Taeg Keun Whangbo. “Automatic salient object extraction based on locally adaptive thresholding to generate tactile graphics”. In : *MDPI Applied Sciences* 10.10 (2020). Publisher: Multidisciplinary Digital Publishing Institute, p. 3350.

[137] Kaiming He, Georgia Gkioxari, Piotr Dollár et Ross Girshick. “Mask R-CNN”. In : *2017 IEEE International Conference on Computer Vision (ICCV)*. ISSN: 2380-7504. Oct. 2017, p. 2980-2988. doi : 10.1109/ICCV.2017.322. url : https://ieeexplore.ieee.org/document/8237584.

[138] Justin Johnson, Andrej Karpathy et Li Fei-Fei. “Densecap: Fully convolutional localization networks for dense captioning”. In : *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, p. 4565-4574. url : http://openaccess.thecvf.com/content_cvpr_2016/html/Johnson_DenseCap_Fully_Convolutional_CVPR_2016_paper.html.

[139] M. Leo, G. Medioni, M. Trivedi, T. Kanade et G. M. Farinella. “Computer vision for assistive technologies”. en. In : *Computer Vision and Image Understanding* 154 (jan. 2017), p. 1-15. ISSN : 1077-3142. DOI : 10.1016/j.cviu.2016.09.001. URL : https://www.sciencedirect.com/science/article/pii/S1077314216301357.

[140] Yue Lu, Chao Guo, Xingyuan Dai et Fei-Yue Wang. “ArtCap: A Dataset for Image Captioning of Fine Art Paintings”. In : *IEEE Transactions on Computational Social Systems* (2022). Conference Name: IEEE Transactions on Computational Social Systems, p. 1-12. ISSN : 2329-924X. DOI : 10.1109/TCSS.2022.3223539.

[141] Marjorie Redon, Lilia Djoussouf, Katerine Romeo, Yvain Quéau et Abderrahim El Moataz. “A Comparison of Audio-Tactile Exploration Methods to Discover the Tapestry of the Apocalypse”. In : *Computers Helping People with Special Needs*. ISSN: 1611-3349. Springer Nature Switzerland, 2024, p. 481-488. ISBN : 978-3-031-62846-7. DOI : 10.1007/978-3-031-62846-7_57. URL : http://dx.doi.org/10.1007/978-3-031-62846-7_57.

[142] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár et Ross Girshick. *Segment Anything*. arXiv:2304.02643 [cs]. Avr. 2023. URL : http://arxiv.org/abs/2304.02643.

[143] Stuardo Lucho, Sylvie Treuillet, Xavier Desquesnes, Remy Leconge et Xavier Brunetaud. “Weakly Supervised SVM-Enhanced SAM Pipeline for Stone-by-Stone Segmentation of the Masonry of the Loire Valley Castles”. In : *Journal of Imaging* 10.6 (2024). Publisher: MDPI, p. 148. URL : https://www.mdpi.com/2313-433X/10/6/148.

[144] Demitrios Galanakis, Stuardo Lucho, Emmanuel Maravelakis, Nikolaos Bolanakis, Antonios Konstantaras, Nectarios Vidakis, Markos Petousis, Sylvie Treuillet, Xavier Desquesnes et Xavier Brunetaud. “Segment Anything Model for Scan-to-Structural Analysis in Cultural Heritage”. In : *2024 5th International Conference in Electronic Engineering, Information Technology & Education (EEITE)*. Mai 2024, p. 1-7. DOI : 10.1109/EEITE61750.2024.10654401. URL : https://ieeexplore.ieee.org/abstract/document/10654401.

[145] N. Parikh, L. Itti, M. Humayun et J. Weiland. “Performance of visually guided tasks using simulated prosthetic vision and saliency-based cues”. eng. In : *Journal of Neural Engineering* 10.2 (avr. 2013), p. 026017. ISSN : 1741-2552. DOI : 10.1088/1741-2560/10/2/026017.

[146] Benoit Deville, Guido Bologna, Michel Vinckenbosch et Thierry Pun. “Guiding the focus of attention of blind people with visual saliency”. In : *Workshop on Computer Vision Applications for the Visually Impaired*. 2008. URL : https://inria.hal.science/inria-00325452/.

[147] Hongchan Yoon, Baek-Hyun Kim, Mukhiddinov Mukhriddin et Jinsoo Cho. “Salient region extraction based on global contrast enhancement and saliency cut for image information recognition of the visually impaired”. In : *KSII Transactions on*

Internet and Information Systems (TIIS) 12.5 (2018). Publisher: Korean Society for Internet Information, p. 2287-2312. URL : https://koreascience.kr/article/JAKO201818564287762.page.

[148] Marwa CHAKROUN, Sonda Ammar BOUHAMED, Imene Khanfir KALLEL, Basel SOLAIMAN et Houda DERBEL. "Indoor Scene Simplification for Safe Navigation Using Saliency Map for the Benefit of Visually Impaired People". en. In : *Advanced Systems for Biomedical Applications*. Sous la dir. d'Olfa KANOUN et Nabil DERBEL. T. 39. Series Title: Smart Sensors, Measurement and Instrumentation. Cham : Springer International Publishing, 2021, p. 201-223. ISBN : 978-3-030-71220-4 978-3-030-71221-1. DOI : 10.1007/978-3-030-71221-1_10. URL : https://link.springer.com/10.1007/978-3-030-71221-1_10.

[149] Karolina PAKĖNAITĖ, Petar NEDELEV, Eirini KAMPEROU, Michael J. PROULX et Peter M. HALL. "Communicating Photograph Content Through Tactile Images to People With Visual Impairments". English. In : *Frontiers in Computer Science* 3 (jan. 2022). Publisher: Frontiers. ISSN : 2624-9898. DOI : 10.3389/fcomp.2021.787735. URL : https://www.frontiersin.org/journals/computer-science/articles/10.3389/fcomp.2021.787735/full.

[150] Karolina PAKENAITE, Eirini KAMPEROU, Michael J PROULX, Adwait SHARMA et Peter HALL. "Pic2Tac: Creating Accessible Tactile Images using Semantic Information from Photographs". en. In : *Proceedings of the Eighteenth International Conference on Tangible, Embedded, and Embodied Interaction*. Cork Ireland : ACM, fév. 2024, p. 1-12. ISBN : 979-8-4007-0402-4. DOI : 10.1145/3623509.3633377. URL : https://dl.acm.org/doi/10.1145/3623509.3633377.

[151] Nian LIU, Junwei HAN et Ming-Hsuan YANG. "PiCANet: Learning Pixel-Wise Contextual Attention for Saliency Detection". en. In : *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, UT : IEEE, juin 2018, p. 3089-3098. ISBN : 978-1-5386-6420-9. DOI : 10.1109/CVPR.2018.00326. URL : https://ieeexplore.ieee.org/document/8578424/.

[152] Bolei ZHOU, Hang ZHAO, Xavier PUIG, Sanja FIDLER, Adela BARRIUSO et Antonio TORRALBA. "Scene parsing through ade20k dataset". In : *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, p. 633-641. URL : http://openaccess.thecvf.com/content_cvpr_2017/html/Zhou_Scene_Parsing_Through_CVPR_2017_paper.html.

[153] Vishnu NAIR, Hanxiu 'Hazel' ZHU et Brian A. SMITH. "ImageAssist: Tools for Enhancing Touchscreen-Based Image Exploration Systems for Blind and Low Vision Users". en. In : *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. arXiv:2302.09124 [cs]. Avr. 2023, p. 1-17. DOI : 10.1145/3544548.3581302. URL : http://arxiv.org/abs/2302.09124.

[154] Daniel OMEIZA, Skyler SPEAKMAN, Celia CINTAS et Komminist WELDERMARIAM. *Smooth Grad-CAM++: An Enhanced Inference Level Visualization Technique for Deep*

Convolutional Neural Network Models. arXiv:1908.01224 [cs]. Août 2019. DOI : 10.48550/arXiv.1908.01224. URL : http://arxiv.org/abs/1908.01224.

[155] Kaiming He, Xiangyu Zhang, Shaoqing Ren et Jian Sun. "Deep Residual Learning for Image Recognition". en. In : *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV, USA : IEEE, juin 2016, p. 770-778. ISBN : 978-1-4673-8851-1. DOI : 10.1109/CVPR.2016.90. URL : http://ieeexplore.ieee.org/document/7780459/.

[156] Michael Dietz, Maha El Garf, Ionut Damian et Elisabeth André. "Exploring Eye-Tracking-Driven Sonification for the Visually Impaired". en. In : *Proceedings of the 7th Augmented Human International Conference 2016*. Geneva Switzerland : ACM, fév. 2016, p. 1-8. ISBN : 978-1-4503-3680-2. DOI : 10.1145/2875194.2875208. URL : https://dl.acm.org/doi/10.1145/2875194.2875208.

[157] Myongchan Kim, Sungkil Lee et Seungmoon Choi. "Saliency-Driven Tactile Effect Authoring for Real-Time Visuotactile Feedback". In : *Haptics: Perception, Devices, Mobility, and Communication*. Sous la dir. de David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu Sudan, Demetri Terzopoulos, Doug Tygar, Moshe Y. Vardi, Gerhard Weikum, Poika Isokoski et Jukka Springare. T. 7282. Series Title: Lecture Notes in Computer Science. Berlin, Heidelberg : Springer Berlin Heidelberg, 2012, p. 258-269. ISBN : 978-3-642-31400-1 978-3-642-31401-8. DOI : 10.1007/978-3-642-31401-8_24. URL : http://link.springer.com/10.1007/978-3-642-31401-8_24.

[158] Laurent Itti et Christof Koch. "Computational modelling of visual attention". In : *Nature reviews neuroscience* 2.3 (2001). Publisher: Nature Publishing Group UK London, p. 194-203. URL : https://idp.nature.com/authorize/casa?redirect_uri=https://www.nature.com/articles/35058500&casa_token=L-rCi1bPiBsAAAAA:fKvN-U_lWh7FtUKLoJ6mQae_TY_bs8W_ZBPZf1_c-TicXnFGgAWY43CdQ60w3jEq4Sv3GYgElAO_ULaKCQ.

[159] Shujie Deng, Julie A. Kirkby, Jian Chang et Jian J. Zhang. "Multimodality with eye tracking and haptics: a new horizon for serious games?" In : *International Journal of Serious Games* 1.4 (2014), p. 17-34. URL : http://eprints.bournemouth.ac.uk/23452/.

[160] Jussi Rantala, Päivi Majaranta, Jari Kangas, Poika Isokoski, Deepak Akkil, Oleg Špakov et Roope Raisamo. "Gaze Interaction With Vibrotactile Feedback: Review and Design Guidelines". en. In : *Human–Computer Interaction* 35.1 (jan. 2020), p. 1-39. ISSN : 0737-0024, 1532-7051. DOI : 10.1080/07370024.2017.1306444. URL : https://www.tandfonline.com/doi/full/10.1080/07370024.2017.1306444.

[161] Susumu Oouchi, Kenji Yamazawa et Lorreta Secchi. "Reproduction of tactile paintings for visual impairments utilized three-dimensional modeling system and the effect of difference in the painting size on tactile perception". In : *Computers Helping People*

with Special Needs: 12th International Conference, ICCHP 2010. Springer Berlin Heidelberg (2010). DOI : 10.1007/978-3-642-14100-3_79.

[162] Kyle Rector, Keith Salmon, Dan Thornton, Neel Joshi et Meredith Ringel Morris. "Eyes-free art: Exploring proxemic audio interfaces for blind and low vision art engagement". In : *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1.3 (2017). Publisher: ACM New York, NY, USA, p. 1-21.

[163] Hyung Nam Kim, Tonya L. Smith-Jackson et Chang S. Nam. "Elicitation of Haptic User Interface Needs of People with Low Vision". In : *International Journal of Human–Computer Interaction* 29.7 (juill. 2013). Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/10447318.2012.722465, p. 488-500. ISSN : 1044-7318. DOI : 10.1080/10447318.2012.722465. URL : https://doi.org/10.1080/10447318.2012.722465.

[164] *Culture et handicap*. fr-FR. Juill. 2024. URL : https://www.culture.gouv.fr/Thematiques/developpement-culturel/Culture-et-handicap (visité le 22/07/2024).

[165] Katerine Romeo, Hannah Thompson et Marion Chottin. "Inclusive Multimodal Discovery of Cultural Heritage: Listen and Touch". en. In : *Computers Helping People with Special Needs*. Sous la dir. de Klaus Miesenberger, Georgios Kouroupetroglou, Katerina Mavrou, Roberto Manduchi, Mario Covarrubias Rodriguez et Petr Penáz. Cham : Springer International Publishing, 2022, p. 278-285. ISBN : 978-3-031-08648-9. DOI : 10.1007/978-3-031-08648-9_32.

[166] Katerine Romeo, Marion Chottin et Hannah Thompson. "Towards an inclusive approach to discovering artworks". In : *Modelling, measurement and control C* 83.1-4 (2022), p. 14-20. URL : https://hal.science/hal-03910184/ (visité le 28/07/2025).

[167] Estelle Costes. "Prise en compte de la perception tactile dans la conception de représentations d'øeuvres d'art pour les personnes en situation de handicap visuel". PhD Thesis. Arts et Métiers Paristech, Paris, 2012.

[168] Beverly Serrell. "Paying Attention: The Duration and Allocation of Visitors' Time in Museum Exhibitions". en. In : *Curator: The Museum Journal* 40.2 (1997). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.2151-6952.1997.tb01292.x, p. 108-125. ISSN : 2151-6952. DOI : 10.1111/j.2151-6952.1997.tb01292.x. URL : https://onlinelibrary.wiley.com/doi/abs/10.1111/j.2151-6952.1997.tb01292.x.

[169] Simon L. Gay, Edwige Pissaloux, Katerine Romeo et Ngoc-Tan Truong. "F2T: A Novel Force-Feedback Haptic Architecture Delivering 2D Data to Visually Impaired People". In : *IEEE Access* 9 (2021). Conference Name: IEEE Access, p. 94901-94911. ISSN : 2169-3536. DOI : 10.1109/ACCESS.2021.3091441.

[170] Hannah Thompson et Vanessa Warne. "Blindness arts: An introduction". In : *Disability Studies Quarterly* 38.3 (2018).

[171] Edwige Pissaloux, Lilia Djoussouf, Katerine Romeo, Velazquez Ramiro, Simon L. Gay, Ngoc-Tan Truong et Son Duy Dao. "A New Approach to Assist Virtual Image Accessibility for Visually Impaired People". en. In : *Universal Access in*

Human-Computer Interaction. User and Context Diversity. Sous la dir. de Margherita ANTONA et Constantine STEPHANIDIS. T. 13309. Series Title: Lecture Notes in Computer Science. Cham : Springer International Publishing, 2022, p. 251-261. ISBN : 978-3-031-05038-18. DOI : 10.1007/978-3-031-05039-8_18. URL : https://link.springer.com/10.1007/978-3-031-05039-8_18.

[172] Simon GAY, Marc-Aurèle RIVIÈRE et Edwige PISSALOUX. "Towards Haptic Surface Devices with Force Feedback for Visually Impaired People". en. In : *Computers Helping People with Special Needs*. Sous la dir. de Klaus MIESENBERGER et Georgios KOUROUPETROGLOU. Cham : Springer International Publishing, 2018, p. 258-266. ISBN : 978-3-319-94274-2. DOI : 10.1007/978-3-319-94274-2_36.

[173] Eduardo FONSECA, Jordi PONS, Xavier FAVORY, Frederic FONT, Dmitry BOGDANOV, Andrés FERRARO, Sergio ORAMAS, Alastair PORTER et Xavier SERRA. "Freesound Datasets: a platform for the creation of open audio datasets". In : *Proceedings of the 18th International Society for Music Information Retrieval Conference (ISMIR 2017)*. Suzhou, China, 2017, p. 486-493.

[174] Stella SYLAIOU et Christos FIDAS. "Supporting People with Visual Impairments in Cultural Heritage: Survey and Future Research Directions". In : *International Journal of Human–Computer Interaction* (2022). Publisher: Taylor & Francis, p. 1-16. DOI : 10.1080/10447318.2022.2098930.

[175] Roberto VAZ, Paula Odete FERNANDES et Ana Cecília ROCHA VEIGA. "Designing an Interactive Exhibitor for Assisting Blind and Visually Impaired Visitors in Tactile Exploration of Original Museum Pieces". en. In : *Procedia Computer Science* 138 (2018), p. 561-570. ISSN : 18770509. DOI : 10.1016/j.procs.2018.10.076. URL : https://linkinghub.elsevier.com/retrieve/pii/S1877050918317228.

[176] Catherine LEROI et Hoëlle CORVEST. *La Tenture de l'Apocalypse d'Angers*. Editions du patrimoine. Sensitinéraires. Déc. 2015. ISBN : 978-2-7577-0312-0.

[177] Salim AKAR, Virginie BEAUCOUSIN, Laetitia VELIN, Charles LENAY, Loïc DESCHAMPS et Vincent ROY. "EXPRESS: Visual-to-auditory sensory substitution with passive movements in a double participant setup". In : *Quarterly Journal of Experimental Psychology* (avr. 2025). DOI : 10.1177/17470218251334990.

[178] Ming JIANG, Shengsheng HUANG, Juanyong DUAN et Qi ZHAO. "Salicon: Saliency in context". In : *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, p. 1072-1080. URL : https://www.cv-foundation.org/openaccess/content_cvpr_2015/html/Jiang_SALICON_Saliency_in_2015_CVPR_paper.html.

[179] Tilke JUDD, Krista EHINGER, Frédo DURAND et Antonio TORRALBA. "Learning to predict where humans look". In : *2009 IEEE 12th international conference on computer vision*. IEEE, 2009, p. 2106-2113. URL : https://ieeexplore.ieee.org/abstract/document/5459462/.

[180] Juan Xu, Ming Jiang, Shuo Wang, Mohan S. Kankanhalli et Qi Zhao. "Predicting human gaze beyond pixels". In : *Journal of vision* 14.1 (2014). Publisher: The Association for Research in Vision and Ophthalmology, p. 28-28. URL : https://jov.arvojournals.org/article.aspx?articleid=2193943&ref=https://githubhelp.com.

[181] Olivier Le Meur, Tugdual Le Pen et Rémi Cozot. "Can we accurately predict where we look at paintings?" en. In : *PLOS ONE* 15.10 (oct. 2020). Sous la dir. de Joseph Najbauer, e0239980. ISSN : 1932-6203. DOI : 10.1371/journal.pone.0239980. URL : https://dx.plos.org/10.1371/journal.pone.0239980 (visité le 23/02/2025).

[182] Mohamed Amine Kerkouri, Marouane Tliba, Aladine Chetouani et Alessandro Bruno. "AVAtt : Art Visual Attention dataset for diverse painting styles". In : *Proceedings of the 2024 Symposium on Eye Tracking Research and Applications*. ETRA '24. New York, NY, USA : Association for Computing Machinery, 2024, p. 1-3. ISBN : 979-8-4007-0607-3. DOI : 10.1145/3649902.3655656. URL : https://doi.org/10.1145/3649902.3655656.

[183] Moran Cerf, E. Paxon Frady et Christof Koch. "Faces and text attract gaze independent of the task: Experimental data and computer model". In : *Journal of vision* 9.12 (2009). Publisher: The Association for Research in Vision and Ophthalmology, p. 10-10. URL : https://jov.arvojournals.org/article.aspx?articleid=2122098.

[184] Markus Bindemann, A. Mike Burton, Ignace T. C. Hooge, Rob Jenkins et Edward H. F. De Haan. "Faces retain attention". en. In : *Psychonomic Bulletin & Review* 12.6 (déc. 2005), p. 1048-1053. ISSN : 1069-9384, 1531-5320. DOI : 10.3758/BF03206442. URL : http://link.springer.com/10.3758/BF03206442.

[185] Markus Bindemann, Christoph Scheepers et A. Mike Burton. "Viewpoint and center of gravity affect eye movements to human faces". In : *Journal of vision* 9.2 (2009). Publisher: The Association for Research in Vision and Ophthalmology, p. 7-7. URL : https://iovs.arvojournals.org/article.aspx?articleid=2193473.

[186] Markus Bindemann, A. Mike Burton, Stephen RH Langton, Stefan R. Schweinberger et Martin J. Doherty. "The control of attention to faces". In : *Journal of vision* 7.10 (2007). Publisher: The Association for Research in Vision and Ophthalmology, p. 15-15. URL : https://jov.arvojournals.org/article.aspx?articleid=2192946&ref=https://githubhelp.com.

[187] S. Rao Jammalamadaka et Ambar Sengupta. *Topics in circular statistics*. T. 5. world scientific, 2001. URL : https://books.google.com/books?hl=fr&lr=&id=6kVqDQAAQBAJ&oi=fnd&pg=PR5&dq=Topics+in+Circular+Statistics&ots=cXQ3tYnyZy&sig=DIhHpQBYOzwKeOCHeBQnxuxyL6E (visité le 01/09/2025).

[188] M. A. Stephens. "Use of the Kolmogorov–Smirnov, Cramér–Von Mises and Related Statistics Without Extensive Tables". In : *Journal of the Royal Statistical Society: Series B (Methodological)* 32.1 (jan. 1970), p. 115-122. ISSN : 0035-9246. DOI : 10.1111/j.2517-6161.1970.tb00821.x. URL : https://doi.org/10.1111/j.2517-6161.1970.tb00821.x.

[189] "Dynamic Time Warping". en. In : *Information Retrieval for Music and Motion*. Berlin, Heidelberg : Springer Berlin Heidelberg, 2007, p. 69-84. ISBN : 978-3-540-74047-6 978-3-540-74048-3. DOI : 10.1007/978-3-540-74048-3_4. URL : http://link.springer.com/10.1007/978-3-540-74048-3_4.

[190] Richard DEWHURST, Marcus NYSTRÖM, Halszka JARODZKA, Tom FOULSHAM, Roger JOHANSSON et Kenneth HOLMQVIST. "It depends on how you look at it: Scanpath comparison in multiple dimensions with MultiMatch, a vector-based approach". en. In : *Behavior Research Methods* 44.4 (déc. 2012), p. 1079-1100. ISSN : 1554-3528. DOI : 10.3758/s13428-012-0212-2. URL : https://doi.org/10.3758/s13428-012-0212-2.

[191] Paul-Christian BÜRKNER. "brms: An R Package for Bayesian Multilevel Models Using Stan". In : *Journal of Statistical Software* 80.1 (août 2017). Section: Articles, p. 1-28. DOI : 10.18637/jss.v080.i01. URL : https://www.jstatsoft.org/index.php/jss/article/view/v080i01.

[192] Fabrice MAUREL, Gaël DIAS, Stéphane FERRARI, Judith-Jeyafreeda ANDREW et Emmanuel GIGUET. "Concurrent Speech Synthesis to Improve Document First Glance for the Blind". In : *2nd International Workshop on Human-Document Interaction associated to ICDAR*. 2019.

[193] Fabrice MAUREL, Stéphane FERRARI, Alexandre BEUDIN et Jean-Marc LECARPENTIER. "Tag Cloud to Tag Thunder: Towards New Oral and Nonvisual Reading Strategies". In : *Actes de la 27ème conférence francophone sur l'Interaction Homme-Machine*. Toulouse, France : ACM, oct. 2015, w19. DOI : 10.1145/2820619.2825021. URL : https://hal.science/hal-01219120.

[194] Dhia Eddine MERZOUGUI, Nilesh TETE, Fabrice MAUREL, Gaël DIAS, Mohammed HASANUZZAMAN, Aurélien BOURNONVILLE, Edgar MADELAINE, Thomas BERTHELIN, Francois LEDOYEN, Laure LEJEUNE-POUTRAIN, François RIOULT et Jeremie PANTIN. "Concurrent Speech and Auditory Tag Clouds for Non-Visual Web Interaction". In : *26th edition of the Interspeech Conference*. Rotterdam, Netherlands, août 2025. URL : https://hal.science/hal-05110274.

[195] Jianxun LOU, Hanhe LIN, David MARSHALL, Dietmar SAUPE et Hantao LIU. "TranSalNet: Towards perceptually relevant visual saliency prediction". In : *Neurocomputing* 494 (2022). Publisher: Elsevier, p. 455-467. URL : https://www.sciencedirect.com/science/article/pii/S0925231222004714.

[196] Marcella CORNIA, Lorenzo BARALDI, Giuseppe SERRA et Rita CUCCHIARA. "Predicting Human Eye Fixations via an LSTM-Based Saliency Attentive Model". In : *IEEE Transactions on Image Processing* 27.10 (oct. 2018). Conference Name: IEEE Transactions on Image Processing, p. 5142-5154. ISSN : 1941-0042. DOI : 10.1109/TIP.2018.2851672. URL : https://ieeexplore.ieee.org/document/8400593.

[197] Dario ZANCA, Stefano MELACCI et Marco GORI. "Gravitational Laws of Focus of Attention". eng. In : *IEEE transactions on pattern analysis and machine intelligence* 42.12 (déc. 2020), p. 2983-2995. ISSN : 1939-3539. DOI : 10.1109/TPAMI.2019.2920636.

[198] Xianyu Chen, Ming Jiang et Qi Zhao. “Predicting Human Scanpaths in Visual Question Answering”. en. In : 2021, p. 10876-10885. URL : https://openaccess.thecvf.com/content/CVPR2021/html/Chen_Predicting_Human_Scanpaths_in_Visual_Question_Answering_CVPR_2021_paper.html.

[199] Mohamed Amine Kerkouri, Marouane Tliba, Aladine Chetouani et Alessandro Bruno. “A domain adaptive deep learning solution for scanpath prediction of paintings”. In : *Proceedings of the 19th International Conference on Content-based Multimedia Indexing*. CBMI ’22. New York, NY, USA : Association for Computing Machinery, oct. 2022, p. 57-63. ISBN : 978-1-4503-9720-9. DOI : 10.1145/3549555.3549597. URL : https://doi.org/10.1145/3549555.3549597.

[200] Yann LeCun, Yoshua Bengio et Geoffrey Hinton. “Deep learning”. en. In : *Nature* 521.7553 (mai 2015). Publisher: Nature Publishing Group, p. 436-444. ISSN : 1476-4687. DOI : 10.1038/nature14539. URL : https://www.nature.com/articles/nature14539.

[201] Sepp Hochreiter et Jürgen Schmidhuber. “Long Short-Term Memory”. In : *Neural Computation* 9.8 (nov. 1997), p. 1735-1780. ISSN : 0899-7667. DOI : 10.1162/neco.1997.9.8.1735. URL : https://ieeexplore.ieee.org/abstract/document/6795963.

[202] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg et Li Fei-Fei. “ImageNet Large Scale Visual Recognition Challenge”. en. In : *International Journal of Computer Vision* 115.3 (déc. 2015), p. 211-252. ISSN : 1573-1405. DOI : 10.1007/s11263-015-0816-y. URL : https://doi.org/10.1007/s11263-015-0816-y.

[203] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov et Liang-Chieh Chen. “Mobilenetv2: Inverted residuals and linear bottlenecks”. In : *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, p. 4510-4520. URL : http://openaccess.thecvf.com/content_cvpr_2018/html/Sandler_MobileNetV2_Inverted_Residuals_CVPR_2018_paper.html.

[204] David Brieber, Marcos Nadal, Helmut Leder et Raphael Rosenberg. “Art in Time and Space: Context Modulates the Relation between Art Experience and Viewing Time”. en. In : *PLOS ONE* 9.6 (2014). Publisher: Public Library of Science, e99019. ISSN : 1932-6203. DOI : 10.1371/journal.pone.0099019. URL : https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0099019.

[205] Francesco Walker, Berno Bucker, Nicola C. Anderson, Daniel Schreij et Jan Theeuwes. “Looking at paintings in the Vincent Van Gogh Museum: Eye movement patterns of children and adults”. en. In : *PLOS ONE* 12.6 (2017). Publisher: Public Library of Science, e0178912. ISSN : 1932-6203. DOI : 10.1371/journal.pone.0178912. URL : https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0178912.

[206] Stephen R. H. Langton, Anna S. Law, A. Mike Burton et Stefan R. Schweinberger. "Attention capture by faces". eng. In : *Cognition* 107.1 (avr. 2008), p. 330-342. ISSN : 0010-0277. DOI : 10.1016/j.cognition.2007.07.012.

[207] Markus Bindemann. "*Scene* and *screen* center bias early eye movements in scene viewing". In : *Vision Research*. Vision Research Reviews 50.23 (nov. 2010), p. 2577-2587. ISSN : 0042-6989. DOI : 10.1016/j.visres.2010.08.016. URL : https://www.sciencedirect.com/science/article/pii/S0042698910004025.

[208] Izabela Krejtz, Agnieszka Szarkowska, Krzysztof Krejtz, Agnieszka Walczak et Andrew Duchowski. "Audio description as an aural guide of children's visual attention: evidence from an eye-tracking study". In : *Proceedings of the Symposium on Eye Tracking Research and Applications*. ETRA '12. New York, NY, USA : Association for Computing Machinery, mars 2012, p. 99-106. ISBN : 978-1-4503-1221-9. DOI : 10.1145/2168556.2168572. URL : https://doi.org/10.1145/2168556.2168572.