Integrating haptic feedback in smart devices : multimodal interfaces and design guidelines

Detjon Brahimaj

Université de Lille

multiTOUCH

L2EP
Laboratoire d'électrotechnique et d'électronique de puissance de Lille

EUROPEAN COMMISSION
Research Executive Agency
Marie Sklodowska-Curie Innovative Training Networks

IRCICA
Research Institute

Integrating Haptic Feedback in Smart Devices: Multimodal Interfaces and Design Guidelines

Thèse de doctorat de l'Université de Lille
préparée à IRCICA

École doctorale n°632 Sciences de l'Ingénierie et des Systèmes (ENGSYS)
Spécialité de doctorat: Génie Electrique (GE)

Thèse présentée et soutenue à Villeneuve d'Ascq, le 30/01/2024, par

Detjon Brahimaj

Jury Composition:

Frédéric Merienne Professor, Arts et Métiers, ParisTech	President - Reviewer
Maud Marchal Professor, Université Rennes, INSA/IRISA	Reviewer
André Mouraux Professor, Université catholique de Louvain, IONS	Examiner
Cagatay Basdogan Professor, Koç University, Istanbul, Turquie	Examiner
Frédéric Giraud Professor, Université de Lille, L2EP	Supervisor
Betty Semail Professor, Université de Lille, L2EP	Co-supervisor
Eric Vezzoli Doctor, Associate Director Razer Inc.	Guest

PhD Thesis

Université de Lille

Thèse de doctorat

multiTOUCH

Intégration de la Rétroaction Haptique dans les Appareils Intelligents : Interfaces Multimodales et Directives de Conception

Thèse de doctorat de l'Université de Lille
préparée à IRCICA

École doctorale n°632 Sciences de l'Ingénierie et des Systèmes (ENGSYS)
Spécialité de doctorat: Génie Electrique (GE)

Thèse présentée et soutenue à Villeneuve d'Ascq, le 30/01/2024, par

Detjon Brahimaj

Composition de Jury:

Frédéric Merienne Professor, Arts et Métiers, ParisTech	President - Rapporteur
Maud Marchal Professor, Université Rennes, INSA/IRISA	Rapporteur
André Mouraux Professor, Université catholique de Louvain, IONS	Examinateur
Cagatay Basdogan Professor, Koç University, Istanbul, Turquie	Examinateur
Frédéric Giraud Professor, Université de Lille, L2EP	Directeur de thèse
Betty Semail Professor, Université de Lille, L2EP	Co-directrice de thèse
Eric Vezzoli Doctor, Associate Director Razer Inc.	Invité

.

.

Disclaimer

This thesis has undergone a thorough review process, which included using an AI tool to correct errors and enhance the overall flow and coherence of the text. The author has carefully reviewed and incorporated the suggested improvements. However, it is important to note that the content of this manuscript is solely the work of the author.

ACKNOWLEDGMENTS

When I started my Ph.D. three years ago, it seemed like a lot of time. However, now, looking backward, I realize how quickly time has passed and how much I have changed during this journey. I would like to thank all the exceptional and talented people who surrounded me during this period and shaped the person I am today in various aspects.

I want to start with my supervisors, Prof. Frederic Giraud and Prof. Betty-Lemair-Semail. Thank you for the trust, confidence, availability, valuable advice, and constructive feedback you have given me. Thanks to the Multitouch project, I had the opportunity to travel and meet the great haptic community, which allowed me to get in touch with extraordinary people.

Thanks to all my colleagues at IRCICA - Angelica, Mondher, Pierre, Yisha, Elis, Geremie - and to Anis, a research engineer who helped me on different occasions with his brilliant problem-solving ability and resilient attitude. Thanks for all the laughter and nice moments we spent together.

I also want to thank all the jury members — Frederic Merienne, Maud Marchal, Andre Mouraux, and Cagatay Basdogan — who honored me by accepting to read and evaluate this work.

Thanks to my mum, Suzana, my dad, Artur, and my sister, Ana, for all the support you gave me during these years. We were 1000 km away, but I felt you close to me every day!

Nora, your support has been vital during these years. With you on my side, I feel I could face everything with no fear. You are my home, my safe place where I can stand and get up when the

world seemed so hard to afford. If I'm here today, it's because you believed in me when I did not, and because you gave me strength when mine was not enough.

To our Yoel, you are my sunshine, reminding me how problems seem small when you look at me with those blue and profound eyes, calling me a super daddy. Thank you for being so special.

Last but not least, to you, Yona. Someone told me that doing a Ph.D. and having a baby at the same time was too hard, but I'm glad I didn't hear them because I would have missed the opportunity to have a wonderful baby like you on my side who fills my heart with love and my eyes with beauty.

Questo dottorato lo devo soprattuto a voi tre: al vostro amore, ai vostri sorrisi e al vostro affetto. Vi amo profondamente!

"J'ai été nourri aux lettres dès mon enfance; et, pour ce qu'on me persuadait que par leur moyen on pouvait acquérir une connaissance claire et assurée de tout ce qui est utile à la vie, j'avais un extrême désir de les apprendre. Mais sitôt que j'eus achevé tout ce cours d'études, au bout duquel on a coutume d'être reçu au rang des doctes, je changeai entièrement d'opinion. Car je me trouvais embarrassé de tant de doutes et d'erreurs, qu'il me semblait n'avoir fait autre profit, en tâchant de m'instruire, sinon que j'avais découvert de plus en plus mon ignorance."

Rene Descartes - Cartesio

"Fatti non foste a viver come bruti, ma per seguir virtute e canoscenza."

Dante Alighieri

ABSTRACT

Integrating Haptic Feedback in Smart Devices: Multimodal Interfaces and Design Guidelines

by Detjon Brahimaj

The growing interest in integrating haptic feedback into commercial products is a direct result of advancements in haptic technology. Notably, the proliferation of smartphones and tablets has led to the integration of haptic modalities for various interfaces.

While extensive research has explored the integration of sensory modalities (visual, auditory, tactile) in passive touch, there is a relative dearth of knowledge regarding bimodality or multimodality in the context of active touch. Emerging technologies, like surface haptics, offer opportunities to investigate various aspects related to sensory integration.

This work provides valuable guidelines for developers, drawing from experimental studies in the realm of active touch. Our initial investigation focuses on the temporal relationship between audio and tactile feedback, revealing a critical 200 ms threshold during sliding interactions on a haptic surface. Moreover, we identify an acceptable audio-tactile delay of 109 ms for click gestures with virtual buttons, emphasizing the need to prohibit or minimize haptic delay to less than 40 ms. A comparative study involving sighted and blind individuals unveils a crucial aspect of inclusion: adhering to synchronization boundaries of the sighted population, relative to virtual buttons, allows for the inclusive design of interfaces accommodating both populations. Additionally, we explore the impact of factors such as stereoscopy and surface deformation on the perception of texture roughness, demonstrating that their presence can alter the perceived roughness of smooth textures by over 20%.

Furthermore, our research explores the potential of using vibration headphones for object localization, revealing a sensitivity of 7° for the haptic modality, 8° for auditory feedback, and 6° for

audio-tactile. This highlights not only the viability of haptic feedback in virtual reality for object localization but also the improvement achieved by reinforcing the sensory experience with audio-tactile stimuli.

RÉSUMÉ

Intégration de la Rétroaction Haptique dans les Appareils Intelligents : Interfaces Multimodales et Directives de Conception

by Detjon Brahimaj

L'intérêt croissant pour l'intégration de la rétroaction haptique dans les produits commerciaux est directement lié aux progrès de la technologie haptique. Notamment, la prolifération des smartphones et des tablettes a conduit à l'intégration de modalités haptiques pour diverses fonctions.

Alors que des recherches approfondies ont exploré l'intégration des modalités sensorielles (visuelle, auditive, tactile) dans le toucher passif, il existe un manque relatif de connaissances en ce qui concerne la bimodalité ou la multimodalité dans le contexte du toucher actif. Les technologies émergentes, telles que l'haptique de surface, offrent des opportunités pour étudier divers aspects liés à l'intégration sensorielle.

Ce travail fournit des lignes directrices précieuses pour les développeurs, tirées d'études expérimentales dans le domaine du toucher actif. Notre première investigation se concentre sur la relation temporelle entre les retours audio et tactiles, révélant un seuil critique de 200 ms lors des interactions de glissement sur une surface haptique. De plus, nous identifions un délai audio-tactile acceptable de 109 ms pour les gestes de clic avec des boutons virtuels, soulignant la nécessité de prohiber ou de minimiser le délai haptique à moins de 40 ms. Une étude comparative impliquant des individus voyants et aveugles dévoile un aspect crucial de l'inclusion : le respect des limites de synchronisation audio-tactile de la population voyante, concerne les boutons virtuels, permet la conception inclusive d'interfaces adaptées aux deux populations. De plus, nous explorons l'impact de facteurs tels que la stéréoscopie et la déformation de surface sur la perception de la rugosité des textures, démontrant que leur présence peut altérer la rugosité perçue des textures lisses de plus de 20%.

En outre, notre recherche explore le potentiel de l'utilisation de casques vibrants pour la localisation d'objets, révélant une sensibilité de 7° pour la modalité haptique, de 8° pour la rétroaction auditive et de 6° pour la rétroaction audio-tactile. Cela met en évidence non seulement la viabilité de la rétroaction haptique en réalité virtuelle pour la localisation d'objets, mais aussi l'amélioration obtenue en renforçant l'expérience sensorielle avec des stimuli audio-tactiles.

LIST OF PUBLICATIONS

D. Brahimaj, F. Giraud, B. Semail, Technological issues with multimodal touch input devices, *Workshop presentation, World Haptics 2021, Jul 2021, Montreal, France*

D. Brahimaj, G. Esposito, A. Mouraux, O. Collignon, F. Giraud, B. Semail, Spatiotemporal detection threshold of audio-tactile delays under conditions of active touch with and without a visual clue, *Poster Presentation, Cognitive Neuroscience Society, Apr 2022, San Fransisco, United States*

D. Brahimaj, F. Berthaut, F. Giraud, B. Semail, Cross-modal interaction of stereoscopy, surface deformation and tactile feedback on the perception of texture roughness in an active touch condition, *Conference paper, IHM'23 - 34e Conférence Internationale Francophone sur l'Interaction Humain-Machine AFIHM, Apr 2023, Troyes, France*

D. Brahimaj, G. Esposito, A. Courtin, F. Giraud, B. Semail, O. Collignon, and A. Mouraux,, Temporal detection threshold of audio-tactile delays under conditions of active touch with and without a visual cue, *Conference paper, IEEE World Haptics Conference (WHC), WHC 2023, Jul 2023, Delft, Netherlands*

D. Brahimaj, M. Ouari, A. Kaci, F. Giraud, C. Giraud-Audine, B. Semail, Temporal Detection Threshold of Audio-Tactile Delays With Virtual Button, *Journal paper, IEEE Transactions on Haptics (ToH), In press*

D. Brahimaj, Mondher Ouari, Anis Kaci, F., Christophe Giraud-Audine, B. Semail, Temporal Detection Threshold of Audio-Tactile Delays With Virtual Button, *Conference paper, IEEE World Haptics Conference (WHC), WHC 2023, Jul 2023, Delft, Netherlands*

D. Brahimaj, E. Vezzoli, F. Giraud, B. Semail, Enhancing Object Localization in VR: Tactile-Based HRTF and Vibration Headphones for Spatial Haptic Feedback, *Journal paper, submitted to IEEE Transactions on Haptics (ToH) on 09/10/23, under review*

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

HCI	Human-Computer Interaction
VR	Virtual Reality
AR	Augmented Reality
SA	Slowly Adapting
FA	Fast Adapting
RTL	Right-to-Left
LTR	Left-to-Right
SOA	Stimulus Onset Asynchrony
TOJ	Temporal Order Judgment
PF	Psychometric Function
SJ	Simultaneity Judgment
IR	Infra-Red
DAC	Data Acquisition Card
V	Visual condition
B	Black screen condition

L	Line condition
M	Mean
STD or SD	Standard Deviation
ANOVA	Analysis Of Variance
EEG	Electroencephalogram
PSS	Point of Subjective Simultaneity
JND	Just Noticeable Difference
HF	Haptic First condition
AF	Audio First condition
IQR	Interquartile Range
VH	Vibrating Headphone
HRTF	Head Related Transfer Function
HMD	Head Mounted Device
VE	Virtual Environment
SDK	Software Development Kit
A	Audio condition
H	Haptic condition
AH	Audio-Haptic condition

GENERAL INTRODUCTION

Context

In our rapidly evolving digital landscape, our interaction with technology and access to information have undergone significant transformations. Among the most revolutionary advancements is the development of multimodal interfaces – such devices aim to enhance human-computer interaction by integrating multiple sensory input modes of communication in a coordinated manner with multimedia system output [1].

Traditionally, the field of human-computer interaction was woven predominantly with visual interfaces, exemplified by text and graphics. Nonetheless, this conventional paradigm often struggled to convey intricate information and fathom the nuances of human intent. In response, the evolution of multimodal interfaces emerged as a potent solution, entwining a diverse tapestry of modalities encompassing text, images, speech, and gestures.

At its core, multimodal interfaces seek to enrich communication and improve user experience by leveraging the strengths of each modality. Text excels in conveying precise information, images and videos unfold concepts visually and elicit emotions, speech recognition facilitates hands-free and natural interactions, while gestures introduce an intuitive and immersive dimension, particularly within virtual environments.

This integration of modalities presents a myriad of opportunities across various domains, from education and healthcare to entertainment and smart technologies. By assimilating and interpreting

insights from multiple sources, machines are better poised to decipher human intent, context, and emotional nuances, culminating in interactions that are both efficient and personalized.

Interestingly, the sense of touch as a sensory modality has gained significant attraction in recent years, propelled by the advancements in tactile devices. Technologies such as haptic surfaces and VR tools show promise for broader integration of haptic feedback in forthcoming devices. Haptic interfaces for interactive touch surfaces have become a new area of research, enabling the ability to conjure tactile sensations on an assortment of devices, spanning smartphones, tablets, household appliances, and automotive interfaces to cite a few.

However, challenges remain in creating a seamless multimodal experience. For instance, synchronizing different stimuli (visual, auditory, tactile) to ensure a perfect user experience can be complicated, especially when lags are introduced due to operating system constraints. While several tactile devices are under research and some have been commercialized, like the XploreTouch device by Hap2U, designing, building, and synchronizing a multimodal haptic device are still open research questions.

To address these challenges and improve the integration of haptic feedback in human-computer interaction (HCI) systems and commercial products, the need for a comprehensive multimodal haptic framework emerges. This framework would accurately synchronize various types of feedback, facilitating seamless implementation and enhancing the overall user experience. As technology continues to advance, the potential for multimodal integration, including haptic feedback, holds promising prospects for shaping the future of HCI.

multiTOUCH project

This work is situated within the framework of the multiTOUCH project[2], an ambitious European initiative funded by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement n° 860114. As an Innovative Training Network (ITN), multiTOUCH operates as a collaborative research training and doctoral program, uniting esteemed universities, research institutions, and industrial partners from five different countries (see Fig. 1).

At the heart of the multiTOUCH project lies a captivating goal - to unravel the complexities of sensory integration in active touch (which requires voluntary and self-generated movement, contrary

to passive touch) and comprehend how the inputs from touching, hearing, and seeing harmoniously blend into a unified and coherent multisensory haptic experience. Furthermore, the project endeavors to investigate how compensatory mechanisms arise when one sense is compromised (touch, vision, or audition), utilizing inputs from other senses to convey or reinforce haptic experiences and information transfer, especially in individuals affected by sensory deficiencies.

In an interdisciplinary approach, the multiTOUCH project seeks to explore the intricate interactions and integration of sensory signals within the brain, comparing individuals with and without sensory deficits. By comprehending these fundamental processes, the project aims to facilitate the design of next-generation multisensory (tactile-visual-auditory) feedback devices and optimize their capabilities based on insights gained from experiments conducted by academic partners.

Central to its mission, multiTOUCH is committed to providing high-level training to a new generation of Early Stage Researchers (ESRs) in the multidisciplinary realm of haptics. The project envisions a dynamic and nurturing environment that fosters creativity, instills an entrepreneurial spirit, and guides ESRs towards flourishing careers. Achieving this vision involves a combination of hands-on training, engaging both academic and industrial researchers, imparting essential academic knowledge and soft-skills, and addressing diverse research challenges across neuroscience, computer science, rehabilitation, human-computer interfaces, multisensory tactile displays, and virtual reality.

By uniting expertise from different fields and promoting collaboration among various institutions, the multiTOUCH project strives to unlock groundbreaking advancements in haptic research and revolutionize the future of multisensory human-computer interactions (HCIs). Through their collective efforts, the project's partners envision a world where technology seamlessly integrates with our senses, enhancing our experiences and enriching the lives of individuals from all walks of life.

In the context of the multiTOUCH project, four institutions were involved: the université de Lille in France, the université Chatolique de Louvain in Belgium, the Italian Institute of technology in Italy and the university Stefan cel Mare in Romania. Moreover, three prominent institutions were involved: one French national company, Hap2U, one multi-national company, Go Touch VR (now Razer), and one hospital, Verbund Katholischer Kliniken-Düsseldorf. Additionally, the project recruited 6 Early Stage Researchers (ESRs) as fellows and allocated 4 experienced researchers across these institutions to undergo specialized and advanced training in various research fields. These training sessions played a crucial role in preparing, inspiring, and equipping the fellows with the necessary skills to pursue the

	Beneficiaries	Field / laboratory	Scientist in-Charge
1	University of Lille	Electrical Engineering	Prof. Frédéric Giraud
2	Université catholique de Louvain	Neuroscience	Prof. André Mouraux
3	Instituto Italiano di Tecnologia	Unit for Visually Impaired People	Prof. Monica Gori
4	University "Stefan cel Mare" of Suceava	Machine Intelligence and Information Visualization Research Laboratory	Prof. Radu-Daniel Vatavu
	Partner Organizations		
5	Universita Degli Studi di Genova	InfoMus	Prof. Gualtiero Volpe
6	GoTouch VR	Haptic software	Dr. Eric Vezzoli
7	Verbund Katholischer Kliniken Düsseldorf	Rehabilitation and assistive technologies	Dr. Mülle
8	Hap2u	Haptic hardware	Dr. Cédrick Chappaz

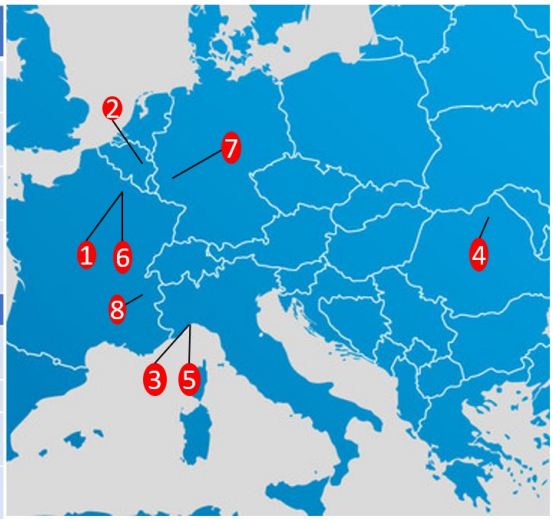

Figure 1: Beneficiaries and partner organizations of the European multiTOUCH project.

project's objectives. Moreover, the training fostered a common communication language and encouraged research collaborations, facilitating exploration of the interdisciplinary aspects of multimodal interactions.

The primary focus of this PhD thesis, carried out within the multiTOUCH project and the L2EP laboratory, centers around the study and development of synchronized multimodal tactile surfaces. Key aspects to be explored include the investigation and characterization of auditory-tactile and visual-tactile integration during active touch condition. To achieve this objective, precise multisensory stimulation is imperative, ensuring tactile feedback is accurately synchronized with auditory and/or visual cues, thereby exploring the temporal dimension of multimodal interactions. The outcomes of the experiments conducted by the project consortium will serve as valuable insights to propose innovative hardware for the next generation of multimodal tactile devices. Building on the experimental results, the research will involve the creation and evaluation of a device designed to incorporate these findings, aiming to optimize its performance in comparison to similar devices in the market. By delving into the intricate world of multimodal interactions and developing cutting-edge technology, this research contributes to a deeper understanding of human-computer interactions. The potential applications of these findings are vast, ranging from enhancing user experiences in virtual reality environments to facilitating more intuitive and engaging human-device interactions. The multiTOUCH project stands

at the forefront of advancing this field, bridging the gap between different institutions and disciplines to shape the future of multimodal interfaces and tactile technologies.

Thesis Structure

The manuscripts is organized as follows:

Chapter 1 comprises an introductory exploration of the human visual and auditory systems, followed by a concentrated investigation into the realm of touch. Within this context, a comprehensive examination of the somatosensory system unfolds, delineating the distinctive contributions of individual mechanoreceptors, their spatial distribution across the body, and their pivotal role in the intricate tapestry of tactile encoding. This discussion further extends to encompass the nuanced realms of tactile perception, including vibrations and textures. The chapter then move its focus towards an in-depth exploration of haptic technologies, from kinesthetic to tactile feedback devices. This survey draws upon a meticulous exploration of the existing literature, delving into the evolution of these technologies aiming to provide comprehensive understanding. Finally, the chapter extends its purview to the multifaceted domain of multimodality and the pivotal role of multimodal integration. Within this broader context, a keen spotlight is cast upon the realms of visuo-haptic and audio-haptic integration.

Chapter 2 delves deeply into the realm of audio-tactile synchronization and underscores its crucial role in multimodal integration. Our exploration of this subject involves a series of meticulously designed experiments aimed at discerning the perceptual detection threshold between auditory and tactile stimuli using a delay injection technique. Our investigation begins by exploring the synchronization in a scenario where participants engage in active touch. Subsequently, we shift our focus to explore perceptual sensitivity in a different context, where participants execute a click gesture on a tactile device. Furthermore, our research extends to the examination of sensitivity disparities between sighted and blind participants. In the end, we provide valuable insights and guidelines for designers and developers in the realm of multimodal haptic interfaces.

Chapter 3 focuses on the exploration of visual-tactile interactions. At first, a bibliographic review into the ways visual modality profoundly shapes our perception is performed with an emphasis on its interactions with 3D depth cues, unraveling the ways in which touch and vision converge to influence our sensory experiences. Furthermore, we shed light on the concept of co-localized devices and their pivotal role in this perceptual interplay and how tactile sensations can manipulate our perception of size, and conversely, how the visual modality can shape our perception of attributes such as softness and roughness. Then, We shift our focus to the captivating interfaces that seamlessly merge the visual and tactile feedback with special attention to the immersive dimensions of stereoscopy found in virtual reality (VR) and augmented reality (AR). To illustrate these concepts in practice, we present an experiment meticulously designed to analyze the influence of visual depth cues, surface deformation, and tactile feedback on the intricate perception of texture roughness during active touch. In the end, design rules and example of application based on our results are outlined.

Chapter 4 concerns the dynamic realm of haptics in conjunction with virtual reality (VR) and our focus sharpens on the delivery of haptic-based directional information, shedding light on various haptic technologies. We start with a bibliographic review where we identify the optimal locations on the body for seamless and effective communication. We then present our interface and perform an experiment where the objective is to evaluate people's accuracy in locating non-visible objects using haptic, audio, or a combination of audio and haptic feedback. This allows us to evaluate user performance and, by extension, assess the spatial resolution of our system, specifically in its capability to deliver precise directional information. We conclude by evaluating the performances and sensitivity of the system illustrating potential and possible applications.

Chapter 5 provides conclusion and prospective of this work.

CHAPTER 1

STATE OF THE ART

Preface to Chapter 1

Human interactions with the external world are inherently multimodal [3]. Through the seamless integration of their senses, humans gain a comprehensive understanding of and adapt to the environment that surrounds them. The senses work in unison continuously and collaboratively, providing valuable insights about the surroundings and enabling meaningful interactions with both the environment and fellow human beings.

Each sense contributes a unique perspective to the human experience. Vision captures the vivid colors, intricate shapes, and dynamic movements of the world. Auditory perception immerses individuals in a rich soundscape, conveying language, music, and environmental cues. Touch offers a tactile connection to the physical world, conveying textures, temperatures, and intimate emotions. Taste and smell further enrich experiences, unlocking the flavors and scents that evoke memories and influence preferences.

The synergy of these senses forms the foundation of how humans perceive, navigate, and interact with their surroundings and multimodal integration enables us to recognize familiar faces, appreciate the nuances of a symphony, and savor the aroma of freshly brewed coffee. It facilitates communication, empathy, and emotional connections with others, enriching our social interactions and shaping our

relationships.

Before delving into the intricacies of each extraordinary sense individually, it is crucial to establish some foundational concepts that apply to all of them. Let us begin by addressing a significant distinction that can often lead to confusion: the difference between sensation and perception [4].

The process by which our sensory organs respond to external stimuli, such as hearing a melody or tasting a delectable dish, is referred to as sensation. It is during these moments that we experience the rain on our head or hear a distant car moving. Sensation involves the engagement of our sense organs in a remarkable phenomenon known as transduction. In this process, physical energy, such as light or sound waves, undergoes a transformation into a form of energy that our brain can comprehend - electrical stimulation. Our sense organs efficiently convert external stimuli into electrical signals that serve as the language of communication within our brains.

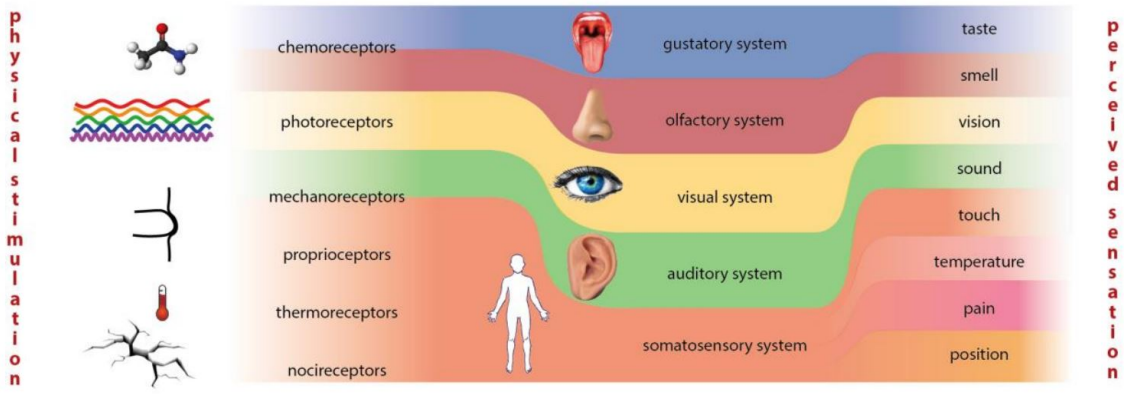

Figure 1.1: Human perception general organization. On the left, we have a physical stimulus that is detected by one or more classes of receptors. In the center, we have the representation of the human sensing system, and on the right, we have the resulting perception derived from the initial stimulation. Adopted from [5]

Once these electrical signals reach our brain, perception begins. This psychological process involves making sense of the vast array of sensory stimulation we encounter, allowing us to appreciate the intricacies of the world around us. During perception, we are endowed with the extraordinary ability to identify a peculiar smell or recognize a familiar melody. A more in-depth discussion on sensation and perception can be found in [4].

It is the intricate interplay between sensation and perception that allows us to navigate our

environment, interpret the sensory input we receive, and construct a coherent understanding of the world. Our senses are the gateway to our experiences, shaping how we interact with the world, form memories, and develop our perceptions of reality. A simplified representation of the process that brings us from sensation to perception is depicted in Fig. 1.1.

As we go through the complexities of multimodal interactions, this research seeks to explore the intricate mechanisms that govern how our senses work together harmoniously or at least to improve our understanding of these interactions. By comprehending the interplay between vision, audition and touch, we aim to uncover the underlying principles that enable us to make sense of the world and unlock new possibilities in fields like human-computer interaction, virtual reality, and assistive technologies. Moreover, by recognizing the significance of sensation and perception, we can enrich our understanding of human cognition and pave the way for innovations in fields such as neuroscience, psychology, and technology.

We will first have an overview on vision and hearing, then we will focus on touch. At first, we will see each sense separately from a biological point of view, moving then into technologies, human perception and multimodal interactions.
In this work, taste and smell are outside our interest and are therefore not considered.

1.1 The sense of sight and hearing

Before exploring human perception and sensory interactions, this section will serve to describe the vision, auditory and tactile system from a biological point of view to have the basis to understand how we perceive the world that surrounds us.

1.1.1 Visual system

The human eye is a complex and remarkable organ that plays a pivotal role in our perception of the world around us. Functionally, the human eye operates much like a camera, capturing light and converting it into electrical signals that our brain processes to create images. When we see an object, we are actually seeing the light that bounces onto the object to our eyes. The light enters the eye through the pupil, a tiny opening behind the cornea. The pupil regulates the amount of light entering the eye by contracting in bright light and dilating in dimmer light. Once past the pupil, light passes through the lens, which focuses an image on a thin layer of cells in the back of the eye called the retina [6].

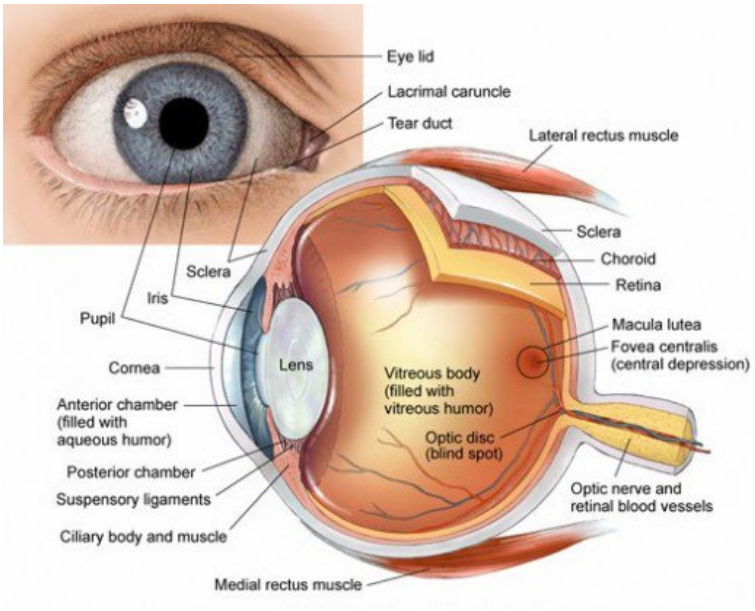

Figure 1.2: Human eye anatomy. Adopted from [7]

Two main photoreceptors are present in the retina: the rods and the cones. Cones provides us with the ability to see colors and fine detail during the day when the light is brighter. On the other hand, rods provide us with the ability to see in a dim light condition, such as during the night. The distribution of these photoreceptors in the retina differs, with rods dominating the periphery and cons highly concentrated in the fovea centralis (see Fig 1.2). The specialized photoreceptors transduce the light entering the pupil into electrical signals that travel down to the optic nerve. From here, the electric signal passes through the thalamus, eventually going into the primary visual cortex where information related to movement and orientation are integrated together [8]. The electric signal is then sent to different areas of the cortex where more complex processing is developed.

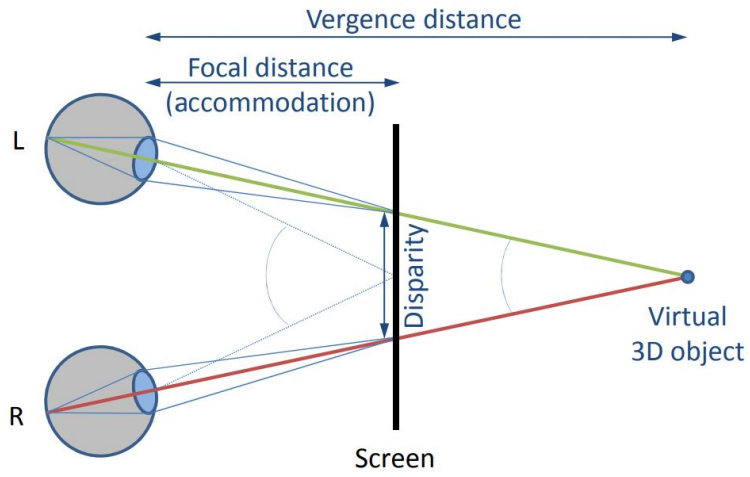

Figure 1.3: Simplified schema of basics for stereoscopy. Adopted from [9]

Human beings possess the remarkable gift of binocular vision, a consequence of having two eyes positioned at slightly different vantage points. This distinctive configuration leads to the phenomenon of binocular disparity, also known as retinal disparity, wherein images focused on the retinas of each eye assume subtly divergent angles. This interocular distinction confers us a remarkable perception of three-dimensional space.

To understand the mechanics of binocular vision, a simplified illustration, as depicted in Figure 1.3, offers elucidation. The core principle entails the projection of a three-dimensional object onto a screen from two distinct positions, effectively simulating the viewpoints of the left and right eyes. The visual perspective of the object's left facet is exclusively discernible to the left eye and analogously for

the right eye. This dichotomy in visual input, encapsulated by the binocular disparity, is seamlessly woven into the brain's fabric as an apparent positional variance, observed along the lines of sight—an optical phenomenon known as parallax [9].

The concept of stereoscopy, rooted in this mechanism, initially expands as a means to provide the illusion of three-dimensionality from a pair of two-dimensional images, merging into what is known as a stereogram. Subsequently, this effect has found widespread application in diverse technological realms. It has been harnessed in a myriad of contexts, spanning Augmented Reality (AR) devices, Virtual Reality (VR) headsets, and 3D television systems, each leveraging distinct technologies to orchestrate this interplay of binocular disparity and human perception.

1.1.2 Auditory system

The auditory system is the organ responsible for the sense of hearing. This system is capable of perceiving sound over a wide range of frequencies ($20\ Hz - 20\ kHz$) known as the audible range, and a wide range of intensities. This system is composed of three main parts: the outer ear, the middle ear and the inner ear, as depicted in Fig. 1.4.

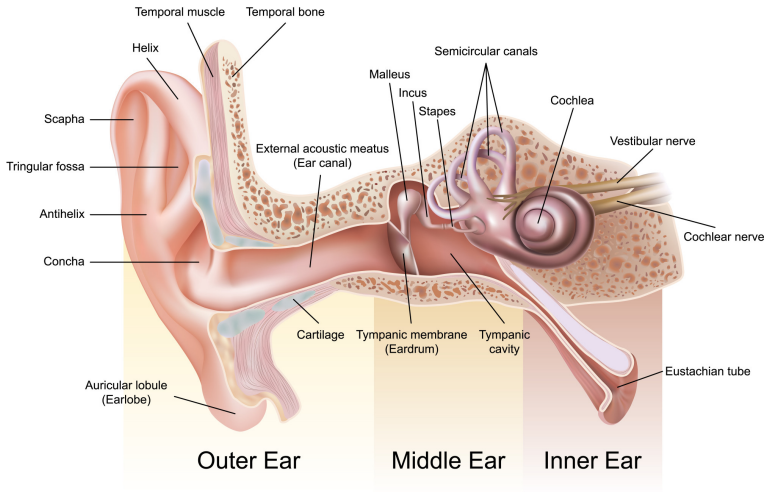

Figure 1.4: Auditory system schematic. Adopted from *Thompsons Road Physiotherapy* [10]

As sound travels from an audio source, such for example a musical instrument, the sound waves are captured by the auricular pinna, which, thanks to its shape, concentrates the waves towards the ear canal in the outer ear. While the auricular pinna amplifies frequencies around $2\ kHz$ by a few

decibels, the ear canal, acting as a resonant pipe, amplifies frequencies in the range $2\ kHz - 5\ kHz$. Indeed, the auditory system is known to be more sensitive in this bandwidth, even if the audible range is larger.

The sound travels from the auricular pinna to the ear canal, reaching the tympanic membrane (eardrum), which has a thickness between $79.6\,\mu$m and $97.0\,\mu$m and a diameter of about $1\,$cm [11]. The sound waves make vibrate the eardrum which converts the acoustic waves into mechanical vibrations. In turn, the eardrum vibrates three tiny bones: the malleus, the incus and the stapes (also called hammer, anvil, and stirrup). These bones amplify the mechanical vibrations and send them to the inner ear. The stape's capability to dump excessive pressure protects the inner ear while the eustachian tube opens periodically to equalize pressure on both sides of the eardrum.

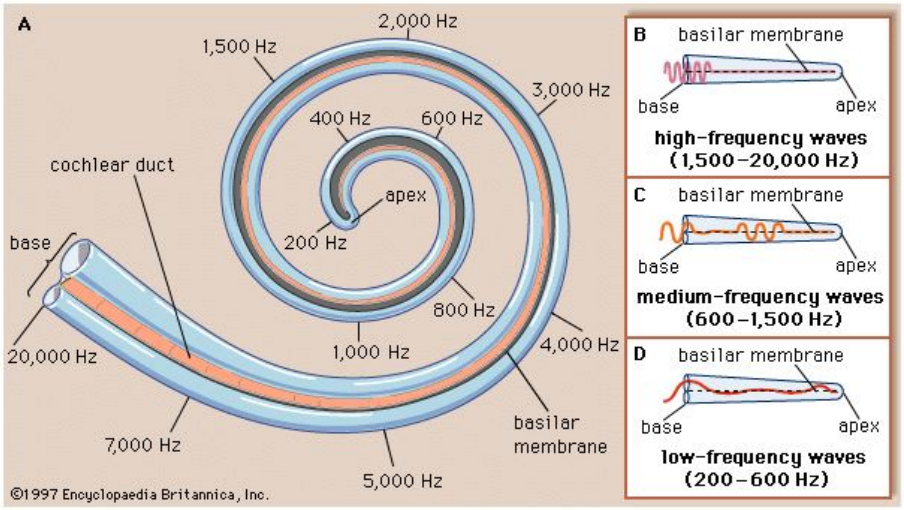

Figure 1.5: Cochlea schematic and acoustic wave propagation. Adopted from *Encyclopedia Britannica* [12]

Vibrations are conveyed from the stapes to the cochlea—a convoluted, fluid-filled hollow structure characterized by three interwoven spiral coiled tubes. Within this intricate architecture, the propagated waves traverse through the fluid, commencing from the base (proximal to the middle ear) and culminating at the apex (the pinnacle or central point of the spiral). This propagation stimulates motion among the inner and outer hair cells, collectively referred to as stereocilia.

The cochlea's intricate spiral arrangement imparts a remarkable capacity for segregating different frequencies to selectively stimulate specific clusters of hair cells dispersed along its convoluted path-

way. This ingenious organization engenders a tonotopic mapping akin to a musical score, wherein distinct pitches find their designated resonance zones, enabling humans to discern and decode the diverse frequency components of sound. This cochlear orchestration and its spatial fidelity are vividly elucidated in Figure 1.5, where the regions responsive to varying frequencies are artistically depicted. This tonotopic sensitivity exhibits a spatial gradient along the cochlea's length. In particular, the base exhibits higher receptivity to higher frequencies while going along the spiral. A progressively greater affinity for lower frequencies becomes discernible, with the apex exhibiting more sensitivity to low frequencies. Crucially, the haircells are responsible for traducing the mechanical vibration into an electrical signal, thanks to the movement of the cilia. The electrical signals are then transmitted to the brain along the cochlear nerve to the primary auditory cortex of the temporal lobe [13].

Remarkably, the presence of a pair of ears on each side of the head offers an innate ability to discern sound's spatial arrangement in three-dimensional space, obviating the necessity of relying solely on our visual senses. While in the visual system a slightly different angulation offers the ability of binocular vision, the time-difference between the two ears offers the ability of sound-source localization or orientation.

1.2 Touch

1.2.1 Overview of somatosensory system

Since the inception of human life, our first and most primal sense is the somatosensory system. Touch is the first sense to develop; for example, it has been reported that tactile responses to a hair stroking the cheek of a fetus around eight weeks gestational age. From then, cutaneous sensitivity of the embryo extends to the genital, palms, soles and abdomen, reaching every part of the body and culminating in a crescendo of responsiveness, where even the faint whisper of a solitary hair's embrace upon the skin becomes palpable at the remarkable milestone of 32 weeks [14]. The somatosensory system serves as a fundamental aspect of our existence, facilitating our ability to thrive and comprehend the world and our own bodies. Without touch, our capacity to skillfully and securely handle objects would be severely hindered. Within the intricate tapestry of the human nervous system, the somatosensory system emerges as a keystone, orchestrating the symphony of perception and

processing across an expanse of sensory modalities, encompassing touch, proprioception, temperature, and pain. This intricate sensory network plays a crucial role in our daily lives, shaping our interactions with the environment and acting as a sentinel guardian for the preservation of our bodily integrity and ultimate survival. Thus, understanding the mechanisms underlying somatosensation is not only of significant scientific interest but also charts a course toward a future where touch converges with technology to reshape human interaction and experience with great potential and versatility across various domains of human knowledge.

Comprised of specialized nerve endings, receptors, and neural pathways, the somatosensory system allows us to perceive external stimuli from our surroundings and internal cues from our own body. The skin, being the largest sensory organ, is populated with an array of mechanoreceptors, thermoreceptors, proprioceptors and nociceptors, each attuned to distinct types of sensory information. These receptors act as transducers, converting mechanical, thermal, and chemical (specifically for the nociceptors) stimuli into electrical signals that are then propagated along a precise neural pathway to the central nervous system.

While many of our senses have specific receptors concentrated in distinct areas of the body, such as the tongue for the gustatory system, the nose for the olfactory system, the eyes for vision, and the ears for audition, the somatosensory system follows a different organizational pattern. The receptors responsible for somatosensationare dispersed throughout the entire body. Their concentrations vary depending on the region of the body under analysis, making the somatosensory system remarkably diverse and adaptable.

Somatosensory information primarily originates from three main sources within the human body. Firstly, from the connective tissues, such as tendons and ligaments, which serve as crucial load-bearing structures. Secondly, from the muscles, responsible for generating motion. And thirdly, from the outermost layer of the body, the skin. Collectively, these sources contribute to our comprehensive understanding of bodily sensations and movements.

Each type of receptor within the somatosensory system serves a distinct purpose in our perception of the world. Mechanoreceptors specialize in detecting mechanical stimuli, such as pressure, vibration, and texture, providing us with essential information about objects' shapes and textures when we touch them. Thermoreceptors, on the other hand, are sensitive to temperature variations, enabling us to perceive both warmth and coldness in our environment. Nociceptors play a crucial role in warning

us of potential harm by transmitting the sensation of pain when tissues are damaged or injured. Finally, proprioceptors serve as an internal navigational system, providing continuous feedback about the position and movement of our body in space.

The dynamic and intricate nature of the somatosensory system allows us to interact with the world in a profound and multifaceted manner. When we grasp a soft, warm object, our mechanoreceptors sense its texture and pressure, while our thermoreceptors perceive its temperature, providing a holistic sensory experience. Similarly, when we encounter a sharp or painful stimulus, our nociceptors trigger a protective reflex, swiftly guiding us away from potential harm.

Understanding the nuances of the somatosensory system unveils the complexity of human perception and our remarkable ability to adapt to diverse sensory environments. Through the integration of information from mechanoreceptors, thermoreceptors, nociceptors, and proprioceptors, we construct a comprehensive understanding of our body and the world around us. However, we have to keep in mind that the exact contributions of the different receptors to the formation of the haptic perception have yet to be established [15].

1.2.2 The skin

In the first place, the skin represents humans as a shield that defends the inner tissue from the external environment and situations such as dehydration, infections, or mechanical stress. The skin is an anisotropic, viscoelastic, non-homogeneous multi-component material [16] composed of different layers.

The skin comprises three distinct layers: the epidermis, the dermis, and the hypodermis. The outermost layer, known as the epidermis, consists of a series of sublayers, with the superficial stratum composed of a layer of deceased cells (keratin) that overlap, measuring roughly $10\,\mu m$ to $40\,\mu m$. Directly beneath the epidermis lies the dermis, a layer replete with a diverse array of specialized sensory neurons. Each neuron within this layer is finely attuned to specific sensory inputs, thereby orchestrating the intricate tapestry of sensations perceived by the human body. This sensory neural network is organized in a specialized manner, as depicted in Figure 1.6. The deepest layer, the hypodermis, assumes a vital role, fulfilling various functions critical to bodily well-being. Among these functions, the hypodermis serves as an energy reservoir, links the dermis to underlying muscles

and bones, provides insulation, and affords protection against external harm. Comprising a complex composition, the hypodermis incorporates elements such as fat, hair follicles, blood vessels, nerves, and sweat glands, collectively contributing to its multifaceted role within the broader skin structure.

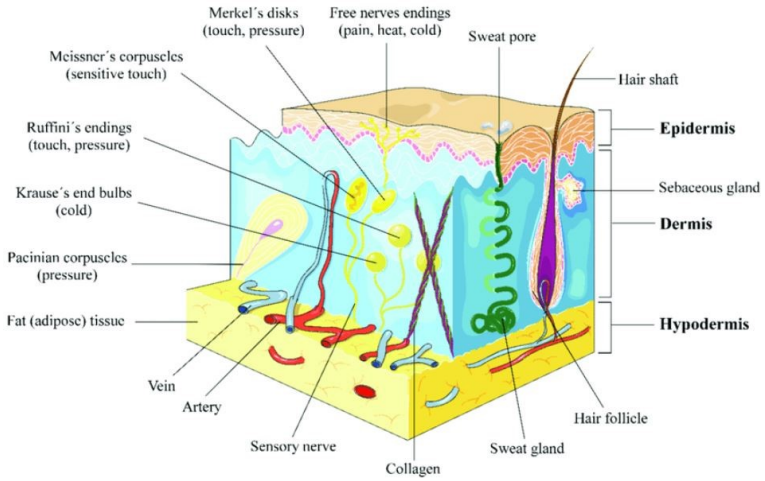

Figure 1.6: Human skin anatomy. Three main layers are highlighted: Epidermis, Dermis and hypodermis. Different receptors are illustrated and each of them is sensitive to specific stimuli such as cold, hot, mechanical vibration, pressure and pain. Adopted from [17]

We have to distinguish three main types of skin, each of which has very different functions and attributes: the mucosal skin, the glabrous skin and the hairy skin. The mucosal skin, being in general humid, covers the internal 'surface' of the body. An example of mucosal skin is the tongue which is capable of detecting a great range of object's attributes such as shape, size, and hardness, to cite a few [18]. It is, therefore, not surprising that babies start to discover objects using their mouths and, in particular, their tongues. Such ability is indeed vital, providing us with the capability to rapidly recognize dangerous objects that can cause mechanical injuries if ingested (such as glass, sand, or splintered bones).

The distinctive attributes of glabrous skin, found in volar regions like the surfaces of the hands and feet, and hairy skin, as exemplified by the dorsal areas of the hand, manifest through four pivotal differentiators: the presence of the pulp, the configuration of ridges, the composition of the stratum corneum, and the distribution of sweat glands. Glabrous skin, situated in proximity to bones, maintains an average distance of approximately $3\,\mathrm{mm}$ to $4\,\mathrm{mm}$ from the underlying skeletal structure. Within

this interstice, a specialized connective tissue known as the pulp resides, endowing the volar hand with its distinctive manipulative and sensory capabilities. This intrinsic pulp presence is in contrast to hairy skin, where such specialized tissue is notably absent. Further distinction arises through the presence of ridges, a characteristic feature exclusive to glabrous skin in regions such as the hand. These ridges play a pivotal role in interactions with objects, effectively reducing the overall contact surface between the skin and external entities [19]. This unique topographical attribute is notably absent in the context of hairy skin. Lastly, the stratum corneum, a layer comprised of deceased cells, engenders a series of intricate mechanical responses during sliding interactions in the glabrous skin [20].

Tactile touch sensation begins at the skin level and in the next section we will see the receptors responsible for tactile touch.

1.2.3 Kinesthetic touch

The term "kinesthesis," initially introduced by [21], originally denoted the perception of movement, a definition that has evolved to encompass the perception of both movement and bodily position. This sensory experience draws upon a diverse array of sensory signals originating from skin receptors, muscles, tendons, and joints [22], collectively contributing to our sense of bodily motion and orientation. This sensation is produced from the interplay of muscle groups alternately contracting and relaxing, or exerting force against one another, thereby altering tensions within the biomechanical system [15]. Specialized mechanoreceptors known as Golgi organs, situated within tendons, provide invaluable feedback to the central nervous system concerning the muscular effort expended in achieving static or dynamic configurations [15]. Complementing this, another variety of mechanoreceptors, the Ruffini corpuscles, resides within joints [23]. These mechanoreceptors respond to deformation, transmitting critical information to the brain regarding limb orientation. Notably, they play a pivotal role, particularly as a joint approaches the limits of its motion range, aiding in the preservation of bodily stability and coordination.

1.2.4 Cutaneous Touch: mechanoreceptors

Within this subsection, we provide an elucidation of distinct mechanoreceptor classes and their specific response, accompanied by a comprehensive exploration of their tactile innervation density throughout the entirety of the human body.

When engaging in the act of touching or grasping an object, the mechanoreceptors embedded within the skin detect mechanical displacements. In the context of glabrous skin, a categorization of four distinct types of mechanoreceptors becomes evident. These types are differentiated based on the corresponding nerve ending's characteristics, encompassing the receptive field (small = type I or large = type II) and by its adaptation speed (Slow-Adapting = SA or Fast-Adapting = FA).

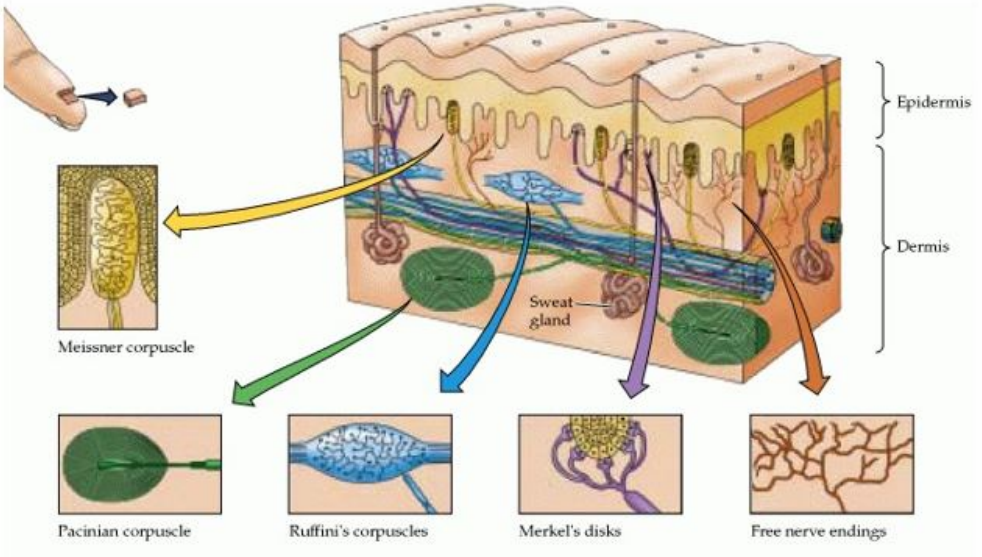

Figure 1.7: (a) Glabrous skin section of a fingertip and the type of mechanoreceptors present in the volar region [24].

The working principle of a mechanoreceptor is based on an osmotic current (firing) produced by the difference in chemical concentration of ions on the two sides of a mechanoreceptor cell's membrane. This asymmetrical charge distribution, triggered by the activation of ionic channels within the mechanoreceptor, subsequently translates into electrical spike potentials (also known as nerve impulses), which are relayed to the brain via the nerve endings. Following the transmission of information and the cessation of tactile stimulation, the ionic concentration is re-established thanks to ionic pumps present in the membrane and the mechanoreceptors are again ready to be activated

by a new stimulation. Below, we elucidate the distinct classes of mechanoreceptors:

- **Meissner cells (FA I)** The type I fast-adapting mechanoreceptors, known as Meissner cells, are strategically positioned in the upper layers of the dermis, maintaining a connection with the epidermis via an intricate network of connective fibers [15]. These receptors exhibit a precisely demarcated receptive field of 3 to 5 mm and show no sensitivity to static skin deformation. Their primary responsiveness lies in dynamic skin deformation, notably displaying efficacy in the face of rapid variations in force (termed transient stimulation). Meissner cells play a pivotal role in activities involving grip and the manipulation of objects. Meissner cells are sensitive to slow skin vibrations between 5 and 40 Hz and their receptive field and adaption is depicted in 1.8.

- **Merkel's disks (SA I)** Positioned at the base of the epidermis, these mechanoreceptors adopt a distinctive organizational structure resembling an intricate tree-like branching system, referred to as a neurite complex [25]. This complex configuration amalgamates 25 to 75 individual receptors into a unique nervous connection. They are sensitive to sustained indentation, encompassing features like points, curvature, or edges, while showcasing a high spatial acuity of approximately 0.5 mm, a notable achievement when compared to their receptive field of 2 to 3 mm. These receptors have traditionally been attributed to form and shape perception; however, recent studies have unveiled their capacity to respond to high-frequency vibrations as well (up to 1500 Hz), similar to Pacinian corpuscle [15].

- **Pacinian corpuscle (FA II)** Dominating in size among the mechanoreceptors, the Pacinian corpuscle is nestled within subcutaneous tissues, occupying the lower stratum of the dermal layer. Its receptive field is broad and its onion-like shape helps the filtering of different frequencies making the pacinian corpuscle sensitive to vibration in the range 40 to 400 Hz with the highest sensitivity at 250 Hz. On texture perception characteristics, the Pacinian and Meissner mechanoreceptors emerge as paramount players, orchestrating the discernment of textural nuances such as variations in roughness.

- **Ruffini endings (SA II)** These mechanoreceptors, present also in joints, are very controversial. Within the realm of cutaneous stimulation, their significance has been deemed marginal due to

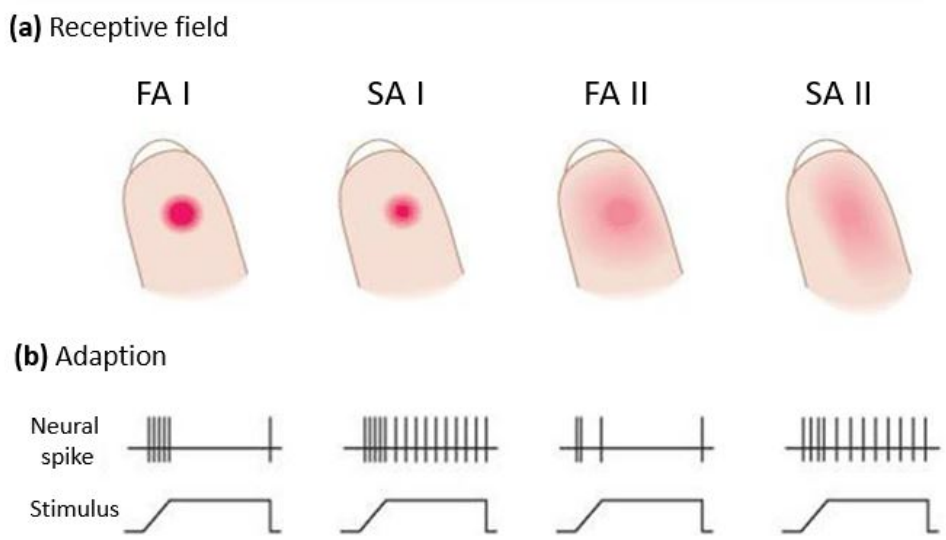

Figure 1.8: (a) Receptive field of each mechanoreceptor classified by adaptation speed. (b) Neural spike train for each mechanoreceptor in response to a stimulus adopted from [26].

their sparse presence within glabrous skin. Instead, they have been primarily associated with the provision of kinesthetic and proprioceptive information [15]. However, their role is unclear due also to the fact that direct observations are rare, and therefore, the debate about their role is still open.

1.2.4.1 Mechanoreceptor density: hand and foot

As elucidated, the categorized mechanoreceptors exhibit distinct functions within tactile perception, and their receptive fields both differ and intersect. The number of tactile afferents (somatosensory neurons) in the glabrous skin of the hand is estimated to be around 17.000 with the number of FA fibers being slightly more than the SA fibers, representing 56% versus 44% respectively. Notably, within the FA, 43% are of type I receptors (Meissner disk) and they are more concentrated in the fingertip while the density decreases by a factor of ten in the palm [27].

Approximately 13% of the FA fibers in the hand belong to the type II classification, characterized by a consistent innervation density across the hand, with a notably higher density observed in the palm region. A quarter (25%) of the SA fibers are categorized as type I, exhibiting a higher density in the fingers while gradually diminishing in the palm area. The remaining 19% consist of SA afferent

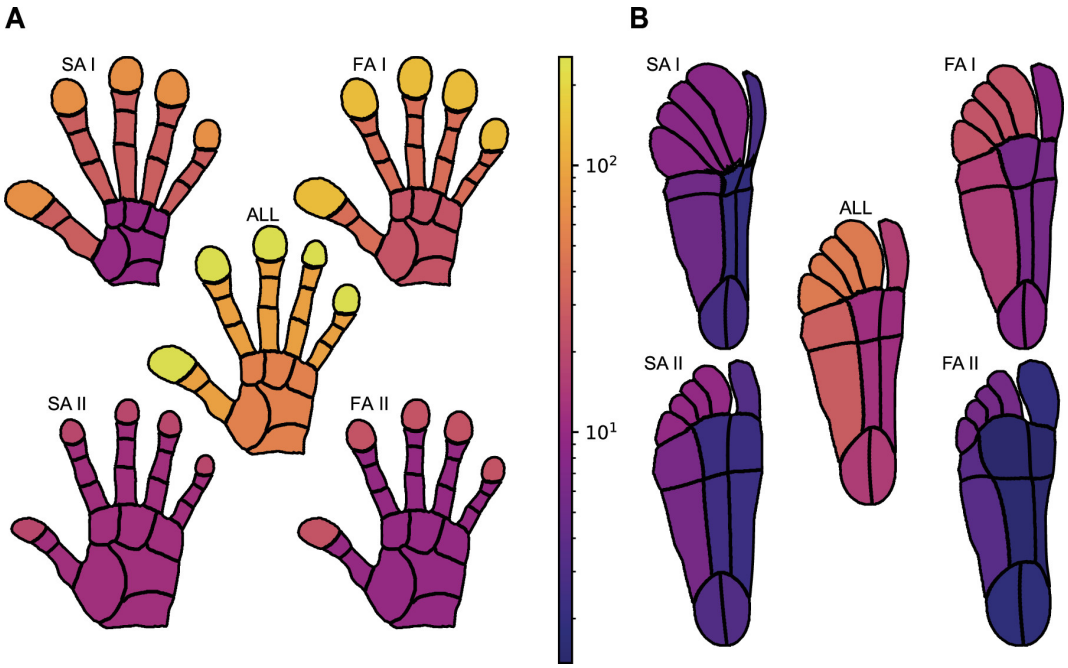

Figure 1.9: Mechanoreceptors' innervation densities in the hand (palmar surface) and foot (plantar surface) in humans. The colors and scale reflect the innervation density (units/cm^2). Adopted from [28].

fibers of type II, which demonstrate a uniform distribution across the glabrous skin of the hand, though certain evidence suggests a relatively heightened density near the skin-nail border [29]. The innervation density of the foot and hand for the four classes of mechanoreceptors is illustrated in Fig. 1.9, employing a flow-based algorithm that preserves border regions while scaling the areas based on a predefined target value, as outlined by [28].

1.2.4.2 Mechanoreceptor density: Whole Body

The regions of the body with the highest density of innervation encompass the palms, hands, fingertips, and face. Additionally, certain areas of the foot, such as the toes, also demonstrate notable innervation.The face is partitioned into three distinct regions.

The initial region, denoted as V1 in Fig. 1.10, encompasses the forehead, eyes, and nose, exhibiting a density of 67 units/cm^2. The second region, labeled as V2 in Fig. 1.10, corresponds to the central facial area, while the third region, designated as V3 in Fig. 1.10, encompasses the lower lip, chin, jaw, and regions around the ears, boasting a density of 84 units/cm^2. The sensitivity of the area around

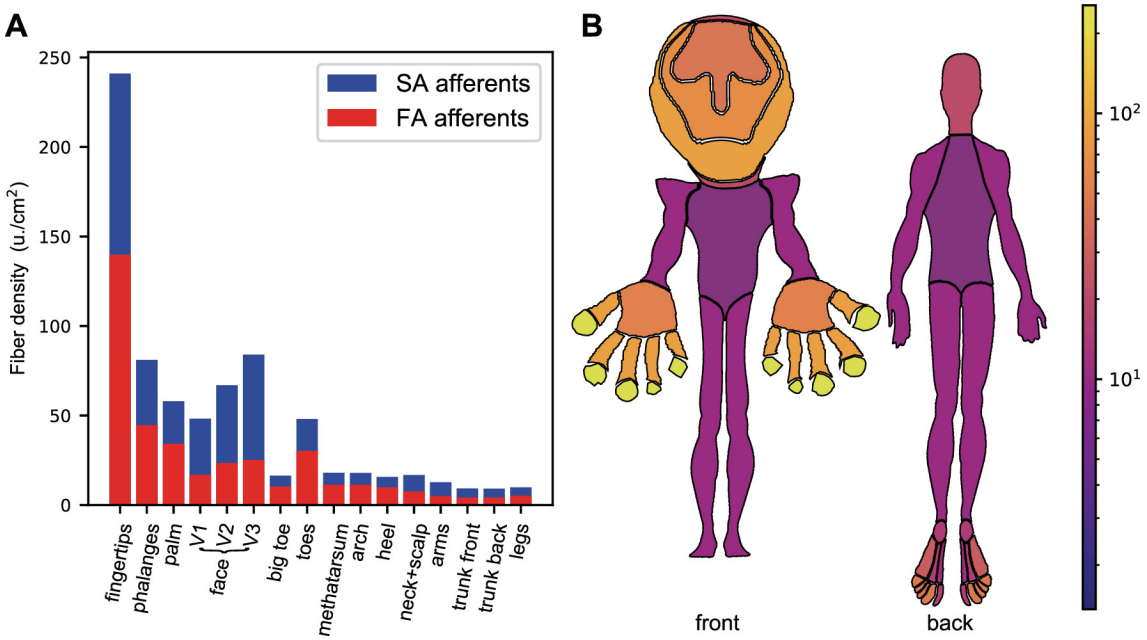

Figure 1.10: Innervation density in the whole body. (A) Innervation density of SA and FA afferents (type I and type II are included). (B) The human body ('homunculi') where colors and scale reflect the innervation density (units/cm^2). Adopted from [28].

the ear is of our interest and will be discussed in Chapter 4. In the remaining part of the body, the least densely innervated areas include the hairy skin of the legs and arms, followed by the trunk.

Drawing upon the variations in innervation density, a diverse array of haptic technologies has emerged over the past decade. In the ensuing section, we will embark on a comprehensive survey of these various haptic technologies. Subsequently, our attention will shift to a detailed exploration of haptic perception.

1.3 Haptic

The term "haptic" encompasses the fusion of two essential perceptual systems [30], namely tactile or cutaneous perception (discussed in Section 1.2.4) and kinesthetic perception (elaborated upon in Section 1.2.3). These systems collectively contribute to the intricate process of actively exploring objects or textures. By synthesizing inputs from the skin, muscles, joints, and tendons, the perceptual experience of haptic allows us to glean information pertaining to weight, shape, hardness, and other attributes of a given object or texture.

Haptic perception can manifest either as passive, where the perceiver lacks control, or active, involving voluntary and self-initiated movements to glean information about surfaces or objects [30]. During the haptic exploration of surfaces or objects, humans employ distinct movement patterns referred to as exploratory strategies to extract specific information or properties. For instance, lateral motion is employed to perceive textures or surface roughness, static contact is used for temperature, and pressure is applied to assess hardness, among other nuanced strategies.

1.3.1 Haptic perception

In the subsequent section, we will delve into a selection of these exploratory strategies, with a primary focus on tactile exploration, as it aligns with the central theme of this study. This exploration aims to provide an insight into the intricate mechanisms underlying our experience of haptic perception.

1.3.1.1 Two-point discrimination

Our perception of vibrations is profoundly influenced by two key factors: the density of sensory receptors in a specific area of the body (with higher concentration leading to greater sensitivity) and the size of the receptors' receptive field (where a smaller field results in higher accuracy). Notably, when two stimuli affect the same receptive field, our body is unable to distinguish between them, allowing us to establish a sensitivity threshold or mean threshold for a two-point discrimination task.

The concept involves gradually bringing two stimuli, designated as two points, closer together until they are perceived as a single entity. This process is illustrated in Fig 1.11c, where we depict the mean threshold across various body locations [31]. Notably, the fingertip displays the highest sensitivity (or lowest threshold), capable of detecting a gap as small as 2mm. Conversely, the upper arm, thigh, and calf demonstrate higher thresholds, making these regions less sensitive in terms of two-point discrimination.

1.3.1.2 Vibration perception

At the vibratory level, mechanoreceptors exhibit sensitivities across a frequency range extending up to 1 kHz. Each mechanoreceptor contributes to the perception of vibratory signals, with its effectiveness

varying depending on the specific frequency involved. Notably, Fast-Adapting (FA) receptors demonstrate heightened sensitivity around frequencies of approximately 200-250 Hz, while Slow-Adapting (SA) receptors surpass the fast-adapting ones at lower frequencies, specifically below 10 Hz, where their sensitivity is more pronounced. Nevertheless, studies have indicated that the smallest vibration amplitudes, as small as 1 µm peak-to-peak at approximately 250 Hz, can trigger firing events [32] (see Fig. 1.11b). In Fig. 1.11a, we present an illustration depicting the involvement of mechanore-

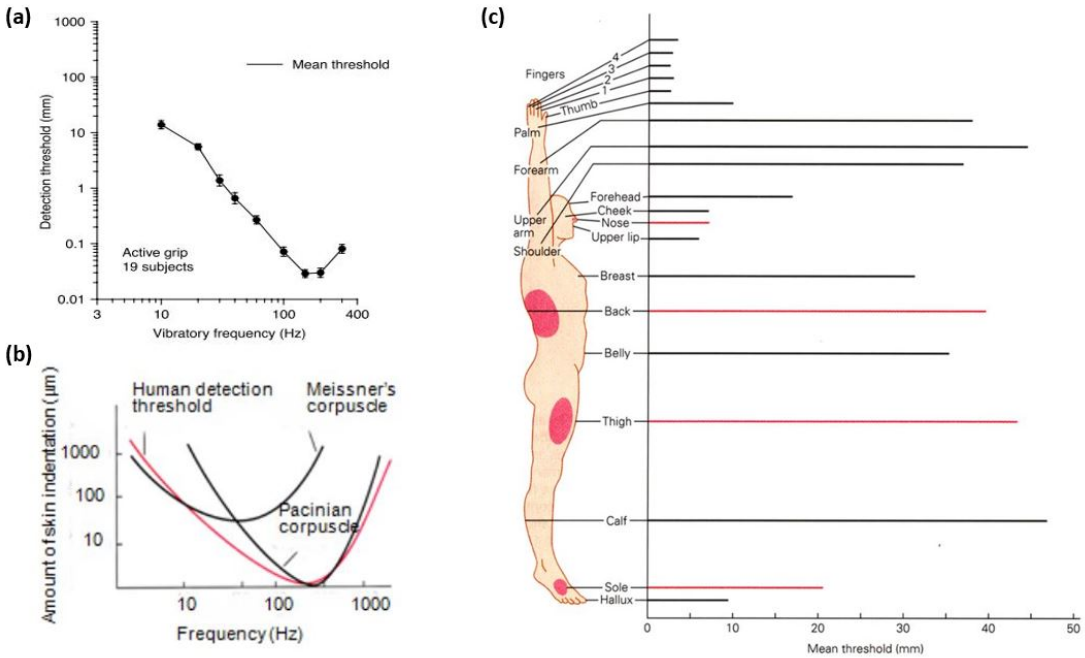

Figure 1.11: (a) Detection threshold when vibrations are transmitted by a 32-mm road cylinder while grasped in the hand by the subjects. Measurements of the vibration amplitude were achieved by a 3-axis accelerometer mounted on the rod. Threshold values and their associated variance are depicted (adopted from [33]). (b) Skin indentation versus frequency. The increasing frequency exhibits a low indentation detection with the highest sensitivity at about 200-250 Hz for approximately 0.06 [32] µm indentation (adopted from [31]). (c) Two points discrimination threshold at different body locations (adopted from [31]).

ceptors in a task involving participants grasping a vibrating cylindrical rod. The represented detection thresholds are in relation to sinusoidal signals. Notably, the human detection threshold aligns with the sensitivity of the Pacinian corpuscle, which manifests a threshold of approximately 250 Hz, as well as the Meissner's corpuscle, with a threshold of around 60 Hz [33]. Consequently, we can attribute a greater contribution to the Fast-Adapting (FA) receptors for the perception of vibrations.

1.3.1.3 Texture perception

The processing of information related to a surface, such as the perception of microgeometry or material properties, is predominantly achieved through tactile exploration. While textural information can also be obtained through vision or audition, touch provides a far more intricate and nuanced understanding of texture. Through touch, we can distinguish between elements that differ in size on the micrometer scale or have spatial periods that vary by hundreds of nanometers [34]. Intriguingly, research has revealed that the scanning velocity of the finger exhibits a high degree of independence from our resulting perception of texture [35], a finding that is remarkable given the variability in exploration velocities observed in tactile tasks [36]. Furthermore, this phenomenon remains notable in light of the fact that tactile receptor responses are influenced by these exploratory parameters [37]. A fascinating phenomenon within the realm of tactile perception arises when large irregularities on a surface, known as asperities, are perceived as distinct events, while finer textures are perceived as continuous. This intriguing observation finds some explanation in the *duplex theory of tactile perception* [38], which posits that the perception of fine texture is intricately linked to the vibrations generated during finger exploration. These vibrations are transmitted to the forearm where the receptive field of FA II receptors is notably expansive. Conversely, the perception of coarse textures is believed to stem from geometric properties that cause skin deformation and trigger the response of SA I receptors. However, recent studies have shown that this mechanism can not be easily divided and is not completely explained by the duplex theory [39].

The perception of textural attributes, such as roughness, has demonstrated its dependence on the inherent properties of the surface. Notably, the sensation of roughness increases as the spatial period decreases, noise is introduced to the texture, or amplitude is heightened [40]. Interestingly, [41] has shown that when textures align well on the epidermal ridge scale, it leads to a higher average intensity in shear force, contributing to friction, and consequently, an increase in resulting vibration upon contact impact. Additionally, investigations have revealed the substantial impact of the visual modality on the perception of roughness [42]. A deeper exploration of this topic will be undertaken in chapter 3, where we investigate the influence of visual cues on roughness perception. Furthermore, an examination of the effects of 3D visual cues on haptic perception will be presented.

1.4 Overview of haptic technologies

As human interactions with the world are incredibly diverse, haptic technologies have been developed to replicate and simulate these interactions. However, due to the vast variety of haptic devices available, classifying them is not a simple assignment.

Culbertson et al. [43] introduced a classification of Haptic Devices based on their interaction methods, dividing them into three major categories: graspable, wearable and touchable. Graspable haptic devices provide kinesthetic feedback, such as force feedback devices. Wearable haptic devices are tactile devices mounted to the hand or other parts of the body. Touchable haptic devices encompass encountered-type displays that enable users to actively explore entire surfaces. Similarly, Wang et al. [44] proposed a classification for haptic devices in virtual reality, categorizing them into 'Desktop haptics' (for virtual tools), 'Wearable haptics' (force feedback gloves, helmets, shoes, etc.), and 'Surface Haptic' (hand-screen or finger-screen interaction).

Additionally, [45] improved this taxonomy focusing on the functionality related to human touch. They classified haptic devices into two main groups: kinesthetic feedback devices involving force feedback and the use of muscles, tendons, and joints, and tactile feedback devices involving skin mechanoreceptors and tactile sensing, such as temperature. Building upon this approach, Basdogan et al. [46] provided a review of surface haptic devices that enable tactile effects on touch surfaces.

Considering the various classifications, groups, and sub-groups, we propose the classification illustrated in Fig. 1.14 for Tactile feedback devices. While providing an overview of the extensive diversity of haptic technologies, our focus will primarily be on tactile feedback devices and, in particular, on surface haptic devices.

1.4.1 Kinesthetic feedback devices

Kinesthetic feedback devices encompass force feedback mechanisms that engage the user's muscles, tendons, and joints, enhancing their interaction experience. These devices are broadly categorized into two main groups: grounded and wearable systems.

1.4.1.1 Force feedback Grounded devices

These devices empower users to engage with and manipulate objects from a distance, spanning across geographical boundaries or within purely virtual environments, achieved through the application of force feedback onto the user's body. Typically, this is accomplished by facilitating user interaction with a fixed, solid body, equipped with a mechanism enabling multi-degree-of-freedom motion.

This remarkable technology establishes a connection between users and task environments, fostering potential enhancements in telepresence. Its versatile applications span various domains, including medical training and teleoperated robotic surgery, among others. At its core, this technology operates on the principle of transmitting counter-forces from a grounded device to the user's hand. The manipulated rigid body can embody virtual tools like surgical scalpels, dental examination mirrors, or dental explorers, to name a few examples. The algorithmic pipeline receives inputs such as force and position, enabling users to experience collision responses (force feedback) that convey tactile sensations of contact forces between virtual objects and avatars. Fig. 1.12 illustrates two examples of grounded devices.

Over the past decade, extensive research has focused on creating realistic prototypes capable of offering six degrees of freedom [44]. Notably, several commercially available devices have emerged, including prominent grounded models such as the TouchX series by 3D Systems Inc. [48], the Quanser Telepresence system [49], and the SPIDAR-G2 (Space Interface Device for Artificial Reality) system [50], among others.

1.4.1.2 Force feedback Wearable devices

Wearable haptic devices present users with the potential for enhanced freedom and expanded workspace, in contrast to their grounded counterparts. These devices are commonly affixed to various body parts, including hands through gloves, legs using braces, and arms as shown in Fig. 1.13b, with possibilities even extending to exoskeleton designs.

Engineered with a focus on virtual reality applications, wearable haptic devices transcend their exclusive association, finding utility in a broader context. This technology provides users with a platform that allows them to move in large virtual environments, encouraging the exploration of intricate gestures for precise manipulations. Notably, these devices are relatively good at producing

Figure 1.12: Examples of grounded haptic devices. Starting from the top: Dentist during training with VOXEL-MAN Virtual Dental Training Simulator [47], TouchX series from 3D System Inc. [48]

sensations on the fingertips or the palm, thereby enriching tactile experiences and facilitating the execution of intuitive gestures. Functioning as portable haptic companions, their versatility extends across an array of domains, encompassing surgical procedures, industrial manufacturing processes, and entertainment realms, among other diverse applications. Fingertip-mounted devices, exemplified by the commercial instance depicted in Fig. 1.13a by GoTouchVR [51], embody a lightweight VR solution. However, their simulation capabilities are primarily geared toward efficient grasping, as the stimulation is concentrated solely on the terminal phalanx of the finger [55]. A notably more versatile approach emerges through the utilization of haptic gloves, a prevalent choice across both industrial applications and research endeavors. These advanced devices transcend the boundaries of kinesthetic feedback, extending their capabilities to encompass a spectrum of tactile responses, achieved through various actuation principles such as pressure, temperature, vibration, and micro-fluid arrays, among other innovative modalities. Two exemplary haptic gloves are showcased in Fig. 1.13: the Manus Quantum Metaglove (Fig. 1.13c), a creation of Manus-Meta [53], and the HaptX glove G1 (Fig.

Figure 1.13: Examples of kinesthetic feedback devices. Starting from the top: Fingertip- mounted devices [51], Space Exoskeleton Controller (SPOC) [52]. On the bottom: Manus Quantum Metaglove [53] and (d) HaptX glove G1 [54].

1.13d), developed by HaptX Inc. [54]; however, a multitude of other commercial alternatives also exist.

1.4.2 Tactile feedback devices

Before going deeper into the diverse technologies underpinning tactile feedback, it is imperative to differentiate between passive and active touch, each encapsulating distinct aspects of tactile experience—receptive touch versus interactive touch. Passive touch involves the reception of tactile stimuli through actuators, without necessitating any user-initiated movement. This form of touch is exemplified by scenarios like a phone's vibration notification, where the user experiences stimuli triggered by external factors (e.g., a received message) independent of their own actions. Conversely, active touch involves deliberate, self-initiated movements, with tactile stimuli arising from the user's engagement

with an object or their exploratory gestures (as seen in technologies such as 1.4.2.4). However, the very act of touching (characterized by active movement) exerts an influence on tactile perception, leading to phenomena like tactile suppression [56] where the ability to perceive tactile stimuli is impaired (i.e. diminished) due to motor activity.

1.4.2.1 Mechanical vibration devices

This category of tactile devices falling within the left branch of Fig. 1.14 represents the prevailing choice for integrating tactile feedback into haptic displays. This preference stems from their cost-effectiveness and seamless integration, presenting an array of options in terms of size, weight, and mechanical functionality. These actuators find application not only in ubiquitous items like smartphones and game controllers but also extend to wearable devices like haptic gloves, as well as a broader spectrum of kinesthetic feedback devices.

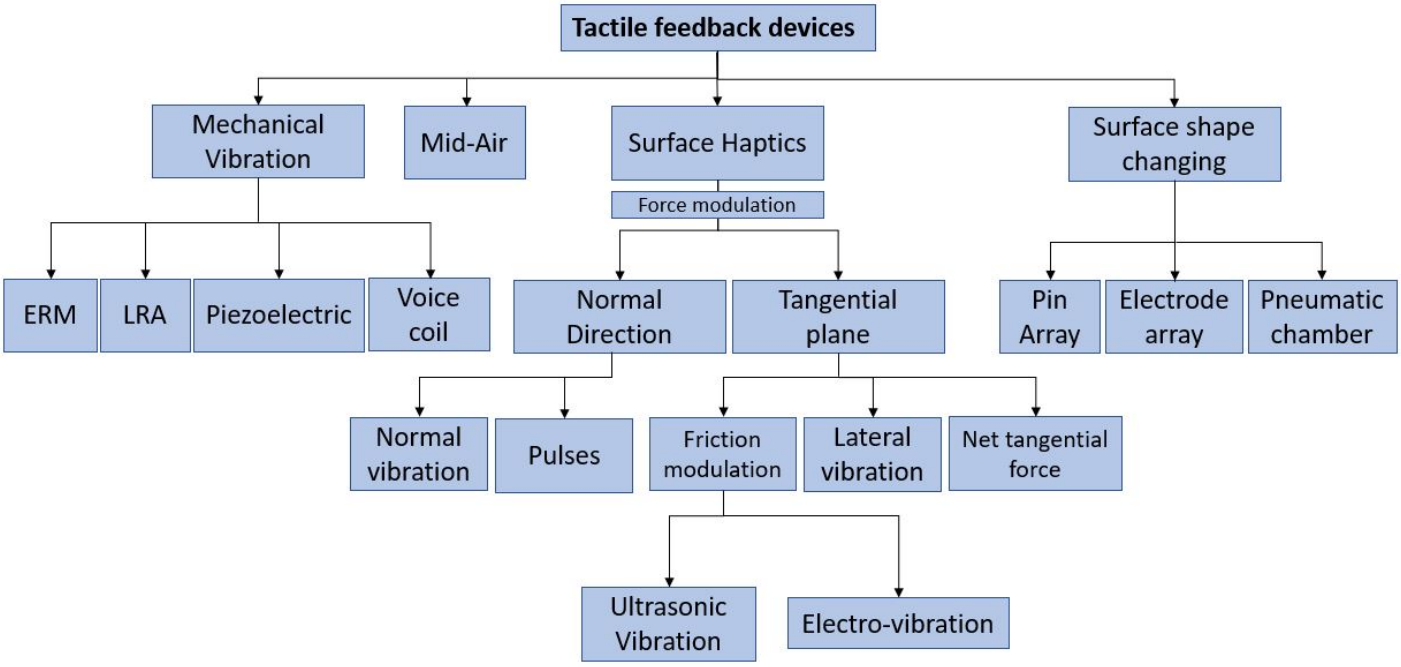

Figure 1.14: Classification of current tactile feedback devices based on the method.

Mechanical vibration actuators serve as proficient tools for generating cutaneous stimuli, allowing for the modulation of tactile sensations through the manipulation of parameters such as frequency,

amplitude, and waveform. This modulation is thoughtfully calibrated to align with the sensitivity of mechanoreceptors. In the selection of a vibrotactile actuator, critical factors encompass not only its physical form but also its responsiveness, power consumption, and input requisites, among other considerations. While comprehensive reviews examine the extensive array of such actuators [57] [58], from a design perspective, the strategic arrangement of these actuators and their placement on the body is pivotal in shaping the resultant perceptual experience. As elucidated in section 1.2.4.2, diverse body locations manifest distinct sensitivities. However, the efficacy of evoking intricate sensations like textures extends beyond just the meticulous placement and selection of actuators. It hinges upon the intricacies of haptic pattern design, encompassing considerations of frequency spectrum and transients. Additionally, predicting the resultant vibrotactile perception is inherently challenging and necessitates a series of design iterations and psychophysical assessments, as highlighted by Wang et al. [44].

1.4.2.2 Mid-air devices

The majority of haptic technologies necessitate a tangible connection between a specific body location and an actuator. Conversely, contactless haptic devices introduce the innovative potential to generate tactile feedback without the requirement for direct physical contact or the utilization of wearable devices.

Various techniques have been cultivated to facilitate tactile feedback through the air. Airstream control is used to employ a controlled air stream to exert pressure and create tactile sensations; sub-woofer compression is used to compress air through apertures, generating tactile effects Thermal effects, induced by lasers and electric arcs, are also employed to provide tactile sensations. Another technique consists on the use of electromagnetic fields that are used in conjunction with magnetic gel or wax applied on the skin to engender tactile sensations. Nonetheless, these methods come with significant drawbacks, including notable time lags and restricted spatial and temporal resolutions. In contrast, ultrasound arrays emerge as a promising solution, effectively mitigating these limitations by providing a comparatively refined tactile feedback mechanism. This advancement enables the faithful rendering of shapes, motion sensations, textured surfaces, and even intricate dynamic patterns in a discernible manner. The implications of this technology have captured the attention of diverse industries, finding applications in sectors such as advertising, automotive advancements, and the realms

of Augmented and Virtual Reality (AR and VR) [59]. Ultrasound arrays have sparked considerable interest owing to their potential to revolutionize tactile experiences in these fields.

First introduced by Iwamoto [60] and then commercialized by Ultraleap Ltd [61], this technique consists of a phase array of ultrasonic transducers used to create a pressurized focal point. The focal point is produced by independently controlling the phase of each emitted wave. Their amplitude peaks are timed to arrive synchronously at a given location, where they constructively interfere to create a focal point with cumulative amplitude (see Fig. 1.15). To generate the focal point, different techniques have been developed. The focal point can be obtained with amplitude modulation (with perceivable frequency, see section 1.2.4), lateral modulation (i.e., the focal point is moved back and forth around the target position) or by spatio-temporal modulation (where the focal point moves continuously along a trajectory to produce skin displacement). Even if this technology is promising, it comes with different limitations such as precision, perception strength, range, size, weight, power consumption and heat dissipation [59].

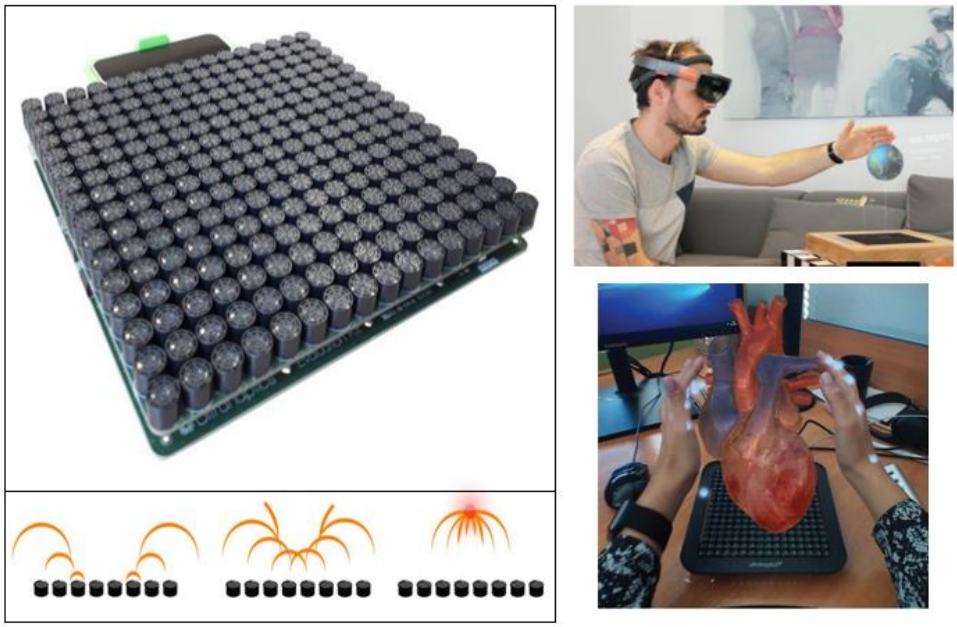

Figure 1.15: On the left there is the Ultralead ultrasonic Mid-Air haptic device and the focal point principle (bottom) [59]. Two examples of mid-air haptic are on the right. Mid-Air Haptic Bio-Holograms in Mixed Reality [62] and Touch Hologram in Mid-Air [63]

1.4.2.3 Surface shape-changing devices

The core principle underlying these particular tactile devices (located within the right branch of Fig. 1.14) revolves around generating tactile sensations by emulating the relief properties of objects. For this purpose, different techniques have been developed to stimulate the user's skin such as the two-dimensional pin array (where the height of the pins is controlled electronically), pneumatic chambers (such as the airbag buttons in Fig 1.16)) and electrode arrays, and electrode arrays that directly stimulate the human nerve to produce a tactile sensation by electrical signals. Surface shape-change devices, while also employed for Braille displays, deliver localized and heightened tactile stimulation, among other haptic devices. However, a potential drawback lies in the dependence on actuator density across the surface, which can present challenges in terms of cost, power consumption and electronic complexity in certain scenarios.

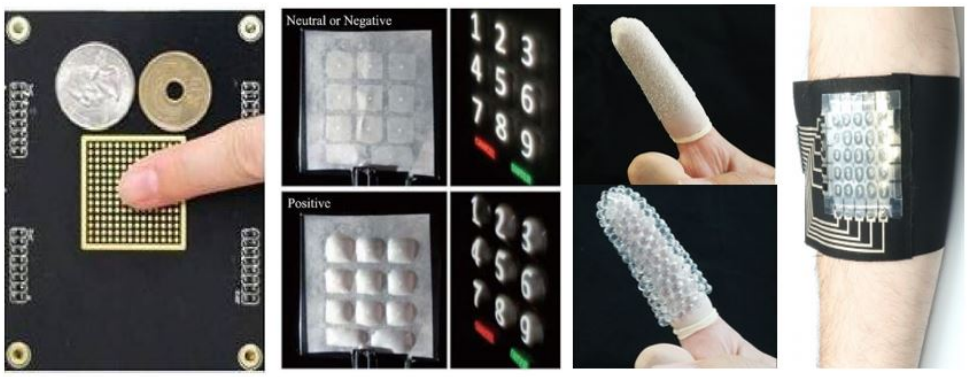

Figure 1.16: Example of surface shape-changing devices. Starting from left: Electrode array display [64], pneumatic airbag buttons [65], multi-point haptic display on the finger [66], Dielectric Elastomer Actuators on the arm [67]

1.4.2.4 Surface Haptics

Haptics for interactive touch surfaces, commonly referred to as surface haptics, represent an emerging frontier within the realm of haptic technology. The fundamental goal of surface haptics revolves around engendering tactile sensations upon touch-sensitive surfaces, encompassing a diverse range of devices like mobile phones, tablets, information displays, and the cutting-edge control panels found in modern home appliances and automobiles. Technologically, at its core, surface haptics endeavors to conceive and engineer innovative devices capable of transmitting tactile feedback to users by

intricately modulating the interaction forces between their fingertips and the touch interface. In a broader context, the existing actuation technologies within surface haptic can be categorized based on the specific direction in which the user's finger experiences stimulation via these interaction forces [46].

Normal direction In cases where force modulation occurs in the normal direction, one prevalent actuation method is vibrotactile, as explored in section 1.4.2.1, where vibrations traverse a screen to reach the user's finger. While this approach proves effective for specific tasks like user notifications, researchers have ventured into more intricate methods to convey nuanced and intricate information, such as localized tactile sensations on specific screen areas. This demand for sophistication has led to the development of various techniques geared towards achieving such localized sensations. Among these techniques are the time reversal method, multimodal decomposition, and phase shift acoustic pulse, among others [46]. In the multimodal decomposition approach, a combination of distinct vibration modes is employed simultaneously. The fundamental concept involves reconstructing an approximation of the desired deformation by determining the contribution from each mode shape, akin to a Fourier series decomposition. Subsequently, the intended localized deformation is applied through the re-composition of each mode's contribution [45]. In the time reversal technique, a localized force is applied on a screen (such as tapping with the finger), which generates propagation of acoustic waves, and this is recorded through sensors applied on the boundaries of the device (screen). The recorded signal is then used as an input for the various actuators. This allows the reconstruction of the initial deformation (finger tap), and consequently, it is possible to reproduce the initial deformation at the precise location where the tap was performed [68].

Lateral direction Lateral vibration can be advantageous due to its reliance on discrete, non-transparent actuators, strategically positioned along the periphery of the designated surface. In this context, electromagnetic and piezoelectric actuators have emerged as prominent choices, attracting considerable attention within the realm of research and development. An alternative avenue for achieving lateral force modulation, free of actuators, involves the indirect generation of lateral force through a method termed "friction modulation." This strategy entails the ability to finely adjust the friction between the fingertip and the interacted surface, achieved through the ingenious utilization

of surface acoustic waves (SAWs) [69]. This manipulation results in tangible tactile sensations and opens the door to the creation of diverse textures that users can perceive [70]. The actualization of friction modulation hinges on the application of two discrete actuation techniques: electrostatic actuation, commonly recognized as electro-vibration, and ultrasonic actuation, often denoted as ultrasonic vibration.

1.4.2.4.1 Friction modulation: Electro-vibration

Discovered by Mallinckrodt et al. [71] during the 1950s, this technique consists of a conductive surface covered by a thin insulating layer. By applying an alternating voltage to the plate and sweeping a dry finger across the surface, a capacitive phenomenon (akin to a parallel plate capacitor) emerges. In this phenomenon, the attractive force between the finger and the plate augments frictional forces, yielding what came to be known as electrovibration (i.e., electrically induced vibration), named by Grimnes et al. [72] which reported that finger humidity could influence the perceived haptic effect. The intricate working principle of this process is illustrated in Fig. 1.17.

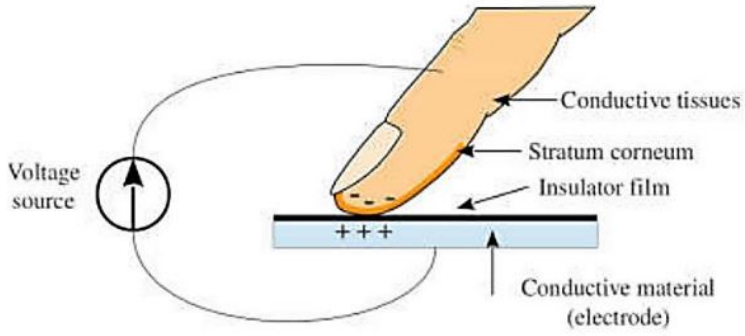

Figure 1.17: Working principle of Electro-adhesion[70]

This innovative principle has underpinned the development of various prototypes, such as the commercial offering Tanvas^{TM} [73], the TeslaTouch [74], and the three-dimensional texture rendering innovation by Disney [75], to name a few prominent instances.

1.4.2.4.2 Friction modulation: Ultrasonic vibration

An ultrasonic wave is a mechanical wave characterized by frequencies surpassing the threshold of human auditory perception (exceeding 20 kHz). These waves exhibit the capacity to traverse diverse

solid materials, encompassing substances such as aluminum and glass. When a finger slides over a surface such as aluminum or glass that is submitted to ultrasonic vibrations, a modification of the friction in accordance with the vibration amplitude of the surface is experienced in the contact area. Achieving ultrasonic vibration on a touch surface often involves gluing piezoelectric actuators to the surface and driving them with sinusoidal voltage signals at the plate's resonance frequency.

The first use of ultrasonic vibration between a surface and a human finger is attributed to Watanabe and Fukui [76] which obtained a surface amplitude of up to 2 μm at a frequency of tens of kilohertz. The authors reported that the friction reduction was increased with increasing the vibration amplitude. Interestingly, when the applied vibration frequency was above 20 kHz, a sensation of softness was perceived on the sandpaper sample they used during their experiment.

Since then, different ultrasonic surface haptic devices have been developed and proposed. Together with the L2EP's StimTac [77] project from the University of Lille (Figure 1.18), the 'T-Pad' project [78] from Northwestern University was one of the first proposals.

Significant research has been carried out at the L2EP laboratory on ultrasonic surface haptic stimulators. From the haptic simulation devices proposed by M. Biet [79], in 2004, the prototype of a tactile simulator [77] was developed. Stimtac prototype has been updated over the years with more sensitive sensors, more robust hardware, and more energy-efficient control. Indeed, a series of 5 small standalone stimtac systems have been developed, using a copper plate for the haptic surface. After that, in 2012, a transparent tactile stimulator was developed [80] (see Fig. 1.18 for the prototype evolution). Finally, Vezzoli et al. [70] developed a tactile feedback tablet, a bimodal co-localized visual-haptic device called E-vita.

Gueorguiev et al. [81] show that human sensitivity to temporary frictional signals, square ultrasonic modulation, implies that the perception of reduced friction in square shapes is influenced by factors such as their sharpness, duration, and the speed of exploration. Consequently, the development of algorithms that precisely control the duration and transition time of such signals plays a pivotal role in perception. On this aspect, the L2EP laboratory has developed algorithms and control strategies such as the Vector Control Method [82].

Ultrasonic friction modulation serves as a versatile tool, extending its capabilities beyond the realm of friction reduction for texture perception. It finds applications in simulating tactile button clicks, a realm where its potential holds promising implications. Despite a notable drawback—namely,

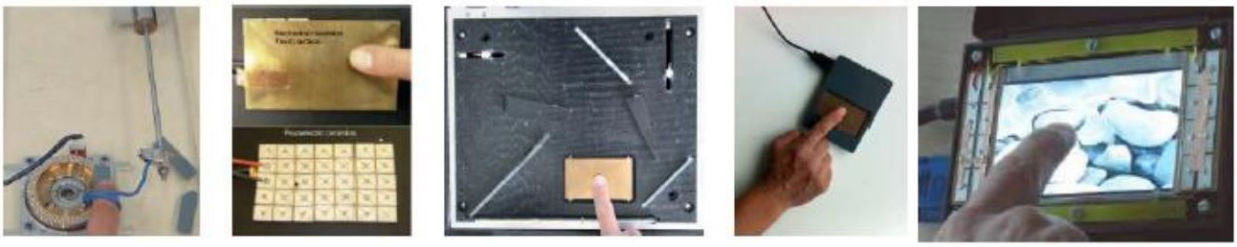

Figure 1.18: Evolution of the StimTac prototype: 1D configuration, 2D feedback, 2D input and feedback (2008), compact US prototype (2010) and transparent prototype (2012)[80]

the generation of audible noise—various researchers have diligently tackled this challenge, unraveling innovative strategies to mitigate its impact. Further exploration into the intricacies of simulating tactile button clicks will be expounded upon in chapter 2, focusing on a more comprehensive analysis of this captivating facet.

The intricate phenomena underlying ultrasonic vibration and the subsequent friction reduction arising from the interaction between the fingertip and the oscillating surface remain veiled in partial understanding. Within the literature, two plausible explanations have been posited: the intermittent contact theory and the squeeze-film effect.

The foundational principle of the squeeze-film effect revolves around the generation of a thin air film undergoing cyclic compression and decompression between the finger-pad and the oscillating surface at high speeds. The idea behind this theory is that the air-film produced by the ultrasonic vibration reduces the contact area between the finger-pad and the surface resulting in a reduction of the frictional forces during sliding. This phenomenon was investigated by Minikes et al. [83] [84], who observed disparate frictional behaviors between a vibrating plate and a non-vibrating one. However, a comprehensive exploration by Sednaoui et al. [85] alludes to the fact that the measured friction reduction cannot be solely attributed to the squeeze-film effect as friction reduction behavior is not fully predicted by it. In their pursuit of a more encompassing explanation to account for the observed data, they advocate the intermittent contact theory, suggesting that the interaction between the finger and the plate is not constant. According to this theory, at specific vibration amplitudes, the finger intermittently loses contact with the oscillating surface, thereby reducing the lateral force exerted by the finger during sliding. This periodic loss of contact culminates in friction reduction between the finger and the oscillating surface.

Although this intermittent contact theory has obtained partial endorsement within the scientific community, a nuanced understanding emerges from recent studies, revealing a collaborative dance between both the squeeze-film effect and intermittent contact as contributors to active lubrication [86]. A proposition advanced by Wiertlewski and Hayward [87] posits that partial squeeze film levitation leads to modulation in perceived friction on the fingertip. However, the precise extent to which each of these mechanisms lends its influence to the phenomenon of active lubrication presents itself as a convoluted enigma, poised for future unraveling and deeper clarification.

1.5 Multimodality

Humans use their senses to understand and comprehend the environment and interactions with the external world are principally multimodal [3]. Indeed, all the senses are used continuously to monitor the environment and, when used collaboratively, they give humans valuable insights about the surroundings and allow them to interact with the environment and other human beings. On the contrary, techniques involved in Human-computer interaction are mainly uni-modal [88], such as the classical keyboard or mouse input methods. Such techniques are effective in different situations but are far from human beings' natural multimodal interaction types. With the rapid advance in non-desktop computing generated by powerful mobile devices and affordable sensors in recent years, multimodal research that leverages speech, touch, and vision is on the rise. Multimodal interaction researchers want to overcome some limitations of the uni-modal condition by adding other modalities, such as touch.

As explained previously, touch is essential in key tasks associated with human survival, as well as many cognitive functions. For this reason, integrating the tactile modality into a multimodal system has innumerable potential applications and could prove beneficial in a very wide variety of areas of human knowledge.

A system that conglomerates multiple input modes provided by a user, such as visual, auditory, or tactile, to achieve a specific output or complete a specific task can be defined as a Multimodal device. According to Oviatt et al. [89], multimodal interfaces process two or more combined user input modes in a coordinated manner with multimedia system output. However, people prefer multiple

input modes over uni-modal input for interaction because it improves the handling, reliability, and personal experience of the user [90]. Although the literature on a formal assessment of multimodal systems is still scattered, several studies have shown that users prefer multimodal interfaces over uni-modal alternatives. These systems can offer better adaptability and reliability by extending interaction alternatives to meet the needs of different users with a range of usage patterns and preferences [91]. As an example, Pitts et al. [92] have developed a multimodal interface for an in-vehicle application with a touchscreen tablet for visual feedback, a speaker for auditory feedback, and an eccentric rotating mass (EMR) actuator for tactile feedback. They evaluated their system in a driving scenario and compared the uni-modal, bi-modal, and multimodal conditions. Their results indicated that, while feedback type did not affect driving or task performance, preference was expressed for multimodal feedback over the other conditions. This result suggests a preference for humans to interact multimodally regardless of performance. Considering this, the increasing interest in multimodal devices seems an inevitable consequence of the need for human beings to live in a multimodal environment.

Nevertheless, different advantages related to task efficiency have been shown in different experiments. Among various advantages, multimodal systems permit the flexible use of input modes, allowing users alternatives in their interaction techniques [93]. Moreover, such systems accommodate a more comprehensive range of users, tasks, and environmental situations because they adapt continuously to changing environmental conditions. Indeed, they accommodate individual differences, such as permanent or temporary handicaps and help prevent overuse of any individual mode during extended computer usage [94]. The author in. [95], for instance, developed a multimodal in-vehicle interactive system using a classical tablet for auditory-visual feedback and an electric motor for tactile feedback. They evaluated their system in a driving scenario while the user had to accomplish a secondary task. As a result, the experimental data indicated that their design could realize eyes-free without influencing the accuracy and completion time of secondary tasks, but also enhancing user experience.

In the realm of multimodal interaction, each distinct modality possesses a signal intricately linked to the intended stimuli. The harmonization of these signals is essential, necessitating meticulous synchronization to seamlessly craft the desired multimodal experience. This intricate orchestration becomes all the more crucial when considering the diverse array of signal types at play. In the domain

of time-sensitive multimodal systems, the task at hand becomes even more intricate. These systems must intricately capture and replicate users' patterns of multimodal integration, accounting for the nuanced shifts in signal timing that manifest across varying circumstances, as highlighted by [96]. Visual scenes, for example, are rendered at a relatively low rate (50 to 120 Hz), audio signals are processed at a high-rate (44.1 kHz), and haptic signals may be rendered even at a higher rate; some tactile signal simulators are computed at frequencies above 60 kHz.

Considering hardware and software, multimodal devices are composed of distinct elements to handle each modality. This makes the design of a multimodal system complex to develop and manage if compared with an uni-modal system. Indeed, different techniques are used to control and manage such systems. Two main approaches can be identified among different methods on how to combine these components and how a multimodal scene can be designed, computed, and rendered. The first one is the Distributed Approach. In this method, complex scenes are distributed into separate computational processes for each modality [97]. However, such an approach poses the problem of defining, mapping and controlling the existing relationships of the processes. Therefore, if the correlation of distinct modalities is not explicit enough, the sensation of realism or believably of a multimodal experience may be of low quality. The second technique, the Single model approach, is based on a single model formalism (one rendering block for all the modalities instead of three). Here, physical modeling of the multimodal scene allows us to create experiences where the object we touch is the one we see and hear. This approach also referred to as multisensory, guarantees coherence between the modalities [98]. However, this approach is typically used for designing multimodal virtual scenes, making it difficult to be used in other scenarios. Fig 1.19 illustrates the block diagram of a multimodal system where visual, auditory, and touch feedback modalities are considered.

1.5.1 Multimodal integration

Multisensory integration, often referred to as multimodal integration, study the intricate interplay of information gleaned from various sensory modalities—such as sight, sound, touch, smell, self-motion, and taste—within the framework of the nervous system [99]. This field explores how the harmonious fusion of modalities constructs a unified representation of objects, affording animals and humans the capacity for rich and meaningful perceptual encounters. Indeed, at the heart of adaptive behavior

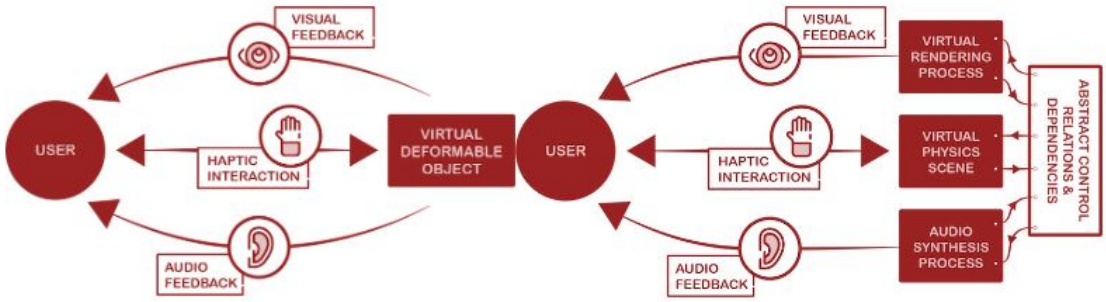

Figure 1.19: Multimodal System Block diagram - Single model approach (left) and Distributed Approach (right). Adapted from [98]

lies multisensory integration, an essential mechanism that bestows the ability to perceive a coherent world replete with perceptual entities. Moreover, this discipline research the intricate dynamics of how diverse sensory modalities interact and influence each other's processing.

A comprehensive glimpse into a multimodal experience, as perceived, is illustrated in Figure 1.20. This experience commences with the activation of distinct receptors — retina for vision, phonoreceptor for auditory, and mechanoreceptors for tactile — each triggered by specific stimuli. Subsequently, the optical, auditory, and tactile stimuli are translated into nerve impulses that traverse our nervous system, ultimately reaching the brain, as elaborated upon in the preceding sections. The juncture where this amalgamation of sensory inputs converges within the brain is labeled as "convergence." It is at this pivotal juncture that the amalgamated information undergoes processing within the brain's realm, constituting the fundamental step of Multisensory processing. Following this, an assessment of the multimodal experience takes place. This evaluation can be stratified into distinct layers [100]. Initially, a multimodal experience can be characterized by a Perceptual layer, which encapsulates the description of stimulus properties—such as categorizing a sound as either loud or soft. This description is subsequently enriched by an Aesthetic layer, where stimuli receive judgments rooted in notions of beauty—whether a sound is deemed pleasant or unpleasant. Furthermore, the experience evaluation encompasses the Meaning layer, where cognition comes into play, attributing expressive qualities or symbolic significance to stimuli—like labeling a sound as elegant. Finally, the Emotional layer contributes its influence, weaving in affective phenomena by assessing whether a sound is surprising, satisfying, or inspiring, thereby culminating in a holistic assessment of the multimodal experience.

As our description unfolds, it becomes evident that this evaluation phase, akin to a rich tapestry,

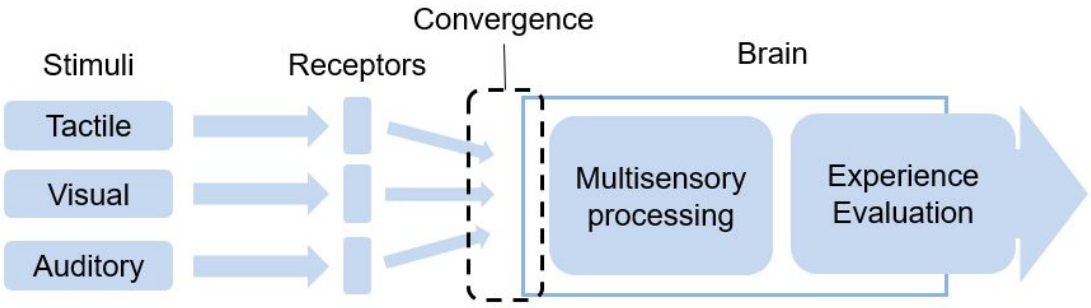

Figure 1.20: Multimodal experience: from Stimuli to Experience Evaluation [100]

encompasses a spectrum of layers. However, for the focus of our exploration, we pivot to the crossroads of visual-haptic and auditory-haptic integration, where the crucible of Multisensory processing, as depicted in Figure 1.20, awaits our attention. To illustrate this interplay, envisage a synchronously perceived burst of sound, a flash, and a vibration triggered by a jolt—all emanating from the same spatial direction. In such an instance, the mind naturally gravitates toward attributing these sensory inputs to a singular source or event. However, it is vital to acknowledge that while the process of multisensory integration is profound, it is not without its imperfections. Occasional missteps transpire during the intricate orchestration of stimuli integration, giving rise to illusions or scenarios where one sense prevails or overcomes the others. An exemplar of such an occurrence is the well-known ventriloquist effect [101], wherein the visual system exerts dominance over the auditory system, resulting in what is termed visual capture. Within this intriguing illusion, the sound originating from the actual speaker appears to emanate from the mouth of a dummy—motioned by the real speaker. This uncanny displacement of sound's origin occurs when spatial congruence between visual and auditory cues is disrupted, leading to the perception of sound emerging from an alternative source (the dummy). The study by Alais and Burr [101] illuminates a nuanced perspective: when visual localization is precise, vision assumes the upper hand in capturing sound; however, when visual stimuli is coupled with noise (blurry or poorly localized visual cues), sound captures the dominion over vision. This underscores the intricate balance at play within the realm of multisensory perception, where the dominance of the visual sense is neither an inviolable rule nor a universal outcome. Rather, instances of multisensory incongruence can lead to a collaborative influence of both modalities upon the final perception.

This collaborative interplay is exemplified in the McGurk effect [102], where an audio-phoneme and a video recording of a distinct phoneme's pronunciation are presented simultaneously. This curious amalgamation of auditory and visual cues results in a compromised perception—a fusion of phonemes that meld into an intermediary form (e.g., 'ba-ba' and 'ga-ga' converging to 'da-da'). When contradictory or incongruent information are received, the brain tends to rely on one of the modality or to combine the information to create a single percept. It is, therefore, interesting to understand which sense will dominate the other and when there is a compromise, which is the contribution of each modality. Interestingly, by investigating size perception with vision and haptic and by introducing noise, Ernst et al. [103] [104] developed a model based on maximum likelihood estimation (MLE) principle, called Bayesian integration, that predicts the multisensory integration outcome. In this model, when a discrepancy between two senses occurs, our brain performs a compromise based on the reliability of each sense. In particular, the dominance of one sense over another is based on the associate variance for each modality; the lower the variance, the higher the reliability. Therefore, when visual dominates over haptic, it is because the variance associated with visual estimation is smaller than that associated with haptic estimation. Moreover, the combined estimate between two senses (visual and haptic, for example) always has a lower variance than the single estimates alone, highlighting that the combination of information from different senses provides a more reliable and precise percept [103]. The theory of Bayesian integration also holds in other situations or with other modalities, such as when exploring a texture with the finger while looking at it [105].

1.5.2 Haptic illusion and Pseudo-haptic effect

Tactile illusions materialize when our tactile perception of a specific attribute of an object deviates from the corresponding physical stimulus detected through the sense of touch [106]. These phenomena span diverse contexts and hold the potential to yield valuable insights into the fundamental mechanisms governing haptic sensations. The prevailing definition of an illusion underscores its essence as a discrepancy between perception and reality. Conversely, the pseudo-haptic effect involves the manipulation of haptic perception by leveraging another sense, typically vision, without the reliance on a physical actuator [107]. It's worth noting that pseudo-haptic effects can emerge even in scenarios where mechanical haptic interfaces are absent [108]. Consequently, many pseudo-haptic

effects in the literature are anchored in visual stimuli that impact the way we perceive distinct haptic attributes such as softness or roughness. In our study (see Chapter 3), we perform an investigation concerning the visual-haptic perception of texture roughness, uncovering a pseudo-haptic effect.

Among the well-known haptic illusions associated with dynamic stimulus sequences, the rabbit illusion is particularly noteworthy. This phenomenon arises when cutaneous stimuli applied to two separate skin locations result in the perception of a tactile stimulus between the chosen spots [109]. Intriguingly, this illusion extends beyond touch and manifests in vision [110] and audition as well. Another compelling illusion is the tau effect, also referred to as perceptual length contraction. It materializes when equally spaced tactile stimuli are perceived unequally based on the timing between the stimuli [111]. Specifically, a shorter temporal interval between stimuli leads to a heightened sense of proximity. Conversely, a greater spatial interval between stimuli induces an illusion of temporal separation—a counterpart to the tau effect known as the kappa effect [112].
Further illusions concerning the interplay with vision will be explored in chapter 3 in section 3.2.1, as this represents a significant aspect of our study's focus.

1.5.3 Visuo-haptic integration

Vision serves as a remarkable tool, actively absorbing intricate spatial information directly through the retina [101]. This perceptual mechanism empowers the brain to construct a detailed and precise comprehension of the environment. The distinctive alignment of vision with a specific realm of representation, notably space, hints at its essential role in forming a high-resolution spatial layout [113]. Over three centuries ago, Berkeley [114] astutely noted that vision lacks immediate access to attributes like distance or 'bigness,' acquiring them only through association with touch. Indeed, Berkeley posited that 'touch calibrates vision,' and recent research indicates that the initial eight years of human life play a pivotal role in brain plasticity and the development of attributes such as binocular vision [115]. However, it's not an absolute truth that touch invariably calibrates vision. Instead, the sense most robustly aligned with a specific task acts as the calibrator [116]. This concept gains traction as studies reveal that while adults seamlessly integrate multimodal data in a statistically optimal manner, children only begin integrating visual-haptic information around the ages of eight to ten [117]. In the domain of spatial perception, haptic dominance emerges in bimodal size perception,

while visual prowess reigns supreme in bimodal orientation perception [118].
Haptic and visuo-haptic object recognition involves the rapid and precise identification of objects. While visual recognition is noted for its speed and accuracy, haptic recognition, although comparably slower, remains remarkably precise, boasting a 96% correct naming rate — 68% within 3 seconds and 94% within 5 seconds [119]. Astonishingly, a mere "haptic glance" of less than 1 second is sufficient in certain cases [120]. The process of haptic identification evolves from an initial "grasp and lift" phase, which collects essential low-level data, to more specialized hand movements such as lateral motions for assessing texture and contour-following for shape analysis [121].

In the realm of categorization, a keystone of higher-order cognition, the spotlight has predominantly focused on visual modality [122]. However, multi-dimensional scaling analyses demonstrated highly congruent perceptual spaces for both modalities concerning novel 3D objects [123]. Significantly, both visual and haptic perceptual spaces adhere closely to the physical object space, preserving category structure [124]. While shape was paramount in visual categorization, haptic and bisensory categorization equally weighed shape and texture [123]. Indeed, cross-modal transfer of category information and structure was evident [125], potentially suggesting a shared multisensory representation.

In texture perception, both haptic and visual system contributes and interact. Heller et al. [126] demonstrated that each modality individually yields a comparable level of performance on roughness perception with the bimodal condition enhancing accuracy. To explore divergent haptic and visual stimuli, Lederman et al. [105] proposed an experimental setup where one half of a texture was visible to subjects, while the other half, hidden behind a curtain, was haptically accessible but not visually discernible. By presenting slightly different half-textures, subjects were asked to match an incongruent standard either with vision, haptic, or both, while the remaining half-texture was placed on a rotating table for rapid stimulus alternation. The result revealed that the bimodal perception emerged as a balanced amalgamation, equally shaped by both visual and haptic, demonstrating that texture exploration can be described by the Bayesian integration theory.

1.5.4 Audio-haptic integration

Within the realm of the senses, hearing emerges as a remarkably precise temporal information captor and communicator [127] [128], pivotal to the construction of multisensory temporal perception. Indeed, our auditory capability extends beyond mere acoustic processing, playing a role in facilitating intricate temporal metric visual representations as well [129] [130]. For instance, studies such as [131] illustrate an intriguing phenomenon: the simultaneous presentation of a flashed spot and two accompanying beeps creates a perception of a dual flash, underscoring an audio-driven influence over visual stimuli. This auditory dominance effect can be influenced by auditory cues presented at varying rates [128]. Moreover, it has been shown with blind subjects that tactile feedback interacts with the auditory spatial localization system, probably due to a process of cross-sensory calibration [132]. In an experiment with blindfolded participants, authors in [133] investigate whether touch is effective in creating a cognitive map of a soundscape and their findings highlight the role of tactile information in modulating auditory spatial tasks, similar to visual cues.

In the realm of texture perception, auditory stimuli play a pivotal role, as evidenced by the ability to discern texture roughness through sound alone [134]. While judgments of roughness through auditory cues alone may fall short of haptic-only assessments, congruent audio-tactile stimuli strengthen performance, aligning closely with haptic-only judgments. Notably, employing a stylus instead of a finger to explore textures results in a roughness estimate that amalgamates the two modalities [135] (weighted: 62% touch, 38% auditory). These findings accentuate the tactile dominance within bimodal integration when it comes to texture perception. However, when the two stimuli are not congruent (i.e., altering sound properties), the tactile perception can be strongly affected, as exemplified by the "Parchment-skin illusion" [136], where altering the audio stimuli affects the tactile perception. To evaluate this effect, authors in [137] performed an experiment where participants judged the relative smoothness or roughness of two abrasive textures with differing grit values. Touch-produced sounds during exploration were recorded and played through earphones while modifying their frequency content in real-time. The results confirmed the "Parchment-skin illusion" and in particular, subjects tended to evaluate more frequently a texture as smooth (or rough) depending on whether the high frequencies were attenuated (or amplified).

The integration of haptic feedback during active touch, characterized by voluntary and self-generated movements, remains less explored in terms of its interactions with other sensory modalities [138]. Previous research has predominantly focused on responses to passive touch [139] [140], underscoring the need to devise experiments that facilitate exploration within the realm of active touch. A significant knowledge gap concerns the perception of temporal synchronization between tactile and auditory stimuli in this context.

Within the framework of multisensory synchronization, an intriguing concept emerges: the brain's vigilant monitoring of temporal coherence across multimodal signals, accommodating potential temporal disparities between senses [141]. The power of multisensory integration appears intricately linked to these temporal relationships among multimodal cues, thereby contributing to heightened performance levels [104]. For this reason, one interest of this work is represented by audio-tactile temporal interaction and, in particular, on sensory synchronization (see chapter 2).

1.6 Conclusion

This chapter commences by delving into the biological underpinnings of our senses, with a specific focus on vision, audition, and touch. Subsequently, a comprehensive categorization of diverse haptic technologies is presented, elucidating the current state of both kinesthetic and tactile feedback devices. The exploration then extends into the realm of multimodality, encompassing topics such as multimodal integration, haptic illusions, and the pseudo-haptic effect. Within this context, we provide detailed insights and examinations into visuo-haptic and audio-haptic integration.

As we draw this chapter to a close, our gaze shifts toward the horizon of audio-tactile integration. The forthcoming chapter will go even deeper, expounding upon the theme of audio-tactile integration by showcasing select investigations we have undertaken within the domain of audio-tactile temporal perception, particularly focusing on sensory synchronization.

CHAPTER 2

AUDIO-TACTILE SYNCHRONIZATION

2.1 Introduction

The seamless fusion of auditory and tactile sensations within multimodal haptic devices has opened the door to a new era of immersive and interactive experiences. Audio-tactile synchronization, the precise coordination of sound and touch feedback, stands as a pivotal element in this technological revolution. It bridges the sensory gap, enhancing our capacity to perceive and interact with digital environments and content in ways that engage not just one but multiple human senses simultaneously. This synchronization unlocks a realm of possibilities, from immersive gaming and virtual reality to accessibility enhancements and advanced simulation systems. In this exploration, we look for the significance of audio-tactile synchronization within multimodal haptic devices, shedding light on perceptual aspects that could play an important role in the sensory integration process.

In this chapter, we will deeply discuss audio-tactile synchronization highlighting its important role in multimodal integration. To investigate this aspect, we designed different experiments to find the perceptual detection threshold between auditory and tactile feedback by using a delay injection technique. At first, we will investigate the synchronicity in a scenario where the participants slide their finger in active touch condition [142]. We will then investigate the perceptual sensitivity in a different scenario, where participants perform a click gesture onto a tactile device [143]. We will

then perform an investigation where we compare sensitivity between sighted and blind individuals and discuss our results, providing guidelines for multimodal haptic designers. This is important not only for inclusive purposes but also to see possible differences between the two populations relative to multimodal integration.

This chapter is adapted from one conference paper [142] done in collaboration with UCL and one journal paper [143].

2.2 Temporal detection threshold of audio-tactile delays under conditions of active touch with and without a visual cue

While much research has been conducted on multisensory interactions in passive touch, research on how active touch influences the interaction between senses remains scarce. Using a haptic surface based on ultrasonic vibrations, we investigated the perception of synchronization of audio-tactile stimuli in active touch. Tactile stimuli were delivered upon sliding the finger, and auditory stimuli were followed with a delay ranging from 0-700 ms. In this simultaneity judgment task, two visual conditions were employed: (i) a visual cue showing the location of the tactile zone on the screen and (ii) a black picture on the screen. We also consider two sliding directions: (i) right-to-left (RTL) and (ii) left-to-right (LTR). Participants slide their finger in the radial to ulnar direction [144].

We estimated the psychometric function (threshold and slope) of the ability to judge whether the auditory and tactile stimuli were temporally synchronous. We found a threshold of 201.26 and 211.73 ms for LTR, and 233.3 and 207.23 ms for RTL, with and without visual cues, respectively. We translated temporal delays into distances (mm) using the finger sliding velocity measured for each trial. The results indicate that the simultaneity judgment was independent of sliding velocity.

2.2.1 Related Work

Sound and touch are natural features of the physical environment that surrounds us. These modalities provide essential information on mechanical impacts and vibrations we experience in everyday life, and all these data are combined by the brain. This process, by which humans merge the available information into a unique perceptual event, is known as multisensory integration [145] and has been

extensively studied, especially for audio-visual interactions. However, how haptic feedback in the condition of active touch, which implies voluntary, self-generated movement [30], integrates with the other senses is more limited [146] since much of the previous research explored responses to passive touch [139]. Therefore, it is essential to design experiments that allow investigation in active touch. In particular, how the temporal synchronization between tactile and auditory stimuli is perceived in this condition remains unclear.

In the context of multisensory synchronization, it has been proposed that the brain monitors temporal coherence among multimodal signals, accounting for possible asynchronies between senses [141]. The advantage of multisensory integration seems closely associated with these temporal relations between the multimodal information and leads to better performances [104].

For example, one of the first investigations that attempted to access temporal perception within auditory and tactile modalities compared people's ability to judge temporal features. The authors [147] presented stimuli either within or across different pairs of sensory modalities and with different stimulus onset asynchronies (SOAs). They used a temporal order judgment (TOJ) task where participants had to judge which stimulus appeared first. Hirsh et al. found that the interval to correctly judge the temporal order was approximately 20 ms for unimodal (touch, auditory, or vision) and multimodal conditions (e.g., audio-tactile).

More recently, Fujisaki et al. [148] investigated the temporal resolution of various modality pairings (audiovisual, audio-tactile, and visual-tactile). Participants were asked to determine if the presented stimuli (single stimulus or repetitive trains) were synchronous or asynchronous. The authors found that audio-tactile pairing was the most reliable for asynchrony detection (15.6 ms) as compared to other sensory pairings, consistent with the study of Hirsh [147].

Other studies have also examined the role of attention in audio-tactile synchronization. For example, in a study by Spence et al. [149], participants were presented with pairs of auditory and tactile stimuli while performing a visual task. The authors found that when the visual task was easy, participants were able to accurately perceive the temporal order of the auditory and tactile stimuli. However, when the visual task was more difficult, their ability to accurately perceive the temporal order of the stimuli was impaired. This result suggests that attentional resources play a role in the

perception of audio-tactile synchronization.

The relative spatial position of stimuli can also modulate the temporal perception of the combined audio-tactile modalities. On this aspect, Ocelli et al. [150] investigated audio-tactile temporal perception with blind and sighted individuals. They compared the co-location of the stimuli against the spatial separation of the stimuli (e.g., stimuli delivered at different locations). Their results demonstrated that the performance of sighted individuals was not affected by the space separation or co-location of the stimuli. At the same time, blind participants were significantly more accurate when the two stimuli were presented from different spatial positions rather than co-located. This result sustains the hypothesis that the lack of a visual cue or visual information, as stimuli presented outside the participant's range of vision [151], is related to more dominant audio-tactile spatial interactions than those arising in the presence of visual information[152]. However, we must consider that blind individuals are not equivalent to sighted individuals temporarily deprived of visual information. Caution is therefore needed when generalizing these results to sighted individuals.

Hence, several characteristics of temporal-perceptual asynchrony have been proposed for some modality pairings. However, the process that governs cross-modal temporal perception is far from being fully understood [153]. At the same time, it represents a critical element of any multimodal feedback system. As the asynchrony between modalities increases, the sense of realism and presence decreases, producing sensory conflicts between modalities and discomfort for the user. Indeed, an appropriate timing relation between modalities is crucial for a congruent perception of multimodal stimuli. Therefore, it is essential to investigate and understand the perceived simultaneity of multimodal stimuli, such as audio-tactile stimuli, especially for active touch.

This work presents the results of an experiment assessing the ability of individuals to detect temporal delays between tactile and auditory stimuli in conditions of active touch. To this aim, we performed a psychophysical experiment in which auditory stimuli were delivered with varying delays relative to tactile stimuli to determine the threshold and slope of the psychometric function (PF) describing the relationship between audio-tactile delay and the probability of detecting the stimuli as asynchronous. Furthermore, we investigated whether a visual cue co-localized with the tactile stimuli had an effect on the ability to discriminate audio-tactile synchronicity.

This investigation is of interest in various fields, including psychology, neuroscience, and engi-

neering, as it can provide insight into how the brain processes sensory information and how different sensory modalities interact. Our results can also help multimodal designers better understand the temporal aspects of audio-tactile interaction and, therefore, design multimodal experiences or devices with a proper timing relation between these modality pairs.

2.2.2 Method

To assess the sensitivity of individuals and evaluate how well participants can detect spatiotemporal delays between pairs of auditory and tactile stimuli, we employed a forced choice method where participants were asked to perform a Simultaneity Judgment (SJ) task. Stimulus placement was continuously optimized using the psi method [154].

The Analysis was performed using R version 4.1.1 and the state-of-the-art platform Stan for Hierarchical model fitting [155]. Stan is a C++ library for Bayesian inference using the No-U-Turn sampler (a variant of Hamiltonian Monte Carlo) or frequentist inference via optimization.

2.2.2.1 Experimental setup

Stimuli were delivered using a visual-tactile display named E-vita. Developed by Vezzoli et al. [156], E-vita (Enhanced Visual-Tactile display) is a flat Haptic Surface based on low-frequency friction modulation through ultrasonic actuation. With this technology, it is possible to modulate the friction between the fingertip and the vibrating plate using acoustic waves thanks to a phenomenon called active lubrication [76] [157]. Transient changes in friction produce a naturalistic tactile input generated by the mechanical interactions between the display and sliding fingertip [79] [158].

E-vita is built around a Banana Pi (Shenzen LeMaker Technology Co. Ltd, China) single-board computer and presents a 1 GHz ARM Cortex-A7 dual-core CPU with 1 GB of RAM and a Lubuntu operating system. The Banana Pi works in parallel with a microcontroller (STM32F4) responsible for the control of the tactile plate. The tactile stimulator comprises 23 piezoceramic components that actuate a $123 * 165 * 2\ mm^3$ fixed glass plate with a resonating sinusoidal mode-shape at $60750\ Hz$, where the half wavelength is 8 mm. 20 piezoceramic components are used as actuators. In contrast, three actuators measure the vibration of the plate provided to the microcontroller. Moreover, an IR (infrared) frame tracks the finger position and velocity, and this information is provided to the

microcontroller, which synthesizes a Pulse Width Modulation signal to pilot a voltage inverter that actuates the piezoceramics. The ultrasonic signal provided to the piezoceramics is modulated at a lower frequency (250Hz) in order to be perceivable by the mechanoreceptors [159]. Furthermore, E-Vita connects to a 7-inch touchscreen, which can display images according to the experiment.

To ensure a low latency in the system's actuation, we coupled E-vita with a Data Acquisition Card (DAC, National Instrument - USB 6363). The tactile stimuli delivered by E-Vita are used as a trigger for the DAC. Once triggered, the DAC can deliver auditory feedback with a maximum latency between the two modalities of around 3.9 μs, which is negligible in our experiment.

2.2.2.2 Stimuli

The stimuli selected for tactile and auditory feedback and the two visual conditions are described hereafter.

Tactile

The tactile stimulus used in this experiment consists of a sinusoidal signal modulated with a square wave with a spatial period of 5000 μm and an amplitude of 40 % (Relative Voltage). These values has been chosen based on preliminary test with the objective to have a easy-to-perceive tactile stimuli with a relatively small amplitude.

The tactile active zone is a rectangle with a height of 100 mm and a width of 2.5 mm centered in the middle of the haptic surface. Hence, tactile feedback was delivered while sliding across the plate when the participant's finger reached the middle of the screen. The co-localized visual-tactile screen is illustrated in Fig. 3.3, and the tactile zone is represented by the black line centered on the screen.

Auditory

Auditory stimulus is delivered through a speaker placed behind the tactile stimulator and in front of the participants. The speaker's location was chosen to maximize the impression that the audio signal came from the interaction of the finger with the device. In this way, as a result, we have a co-localized audio-tactile stimuli.

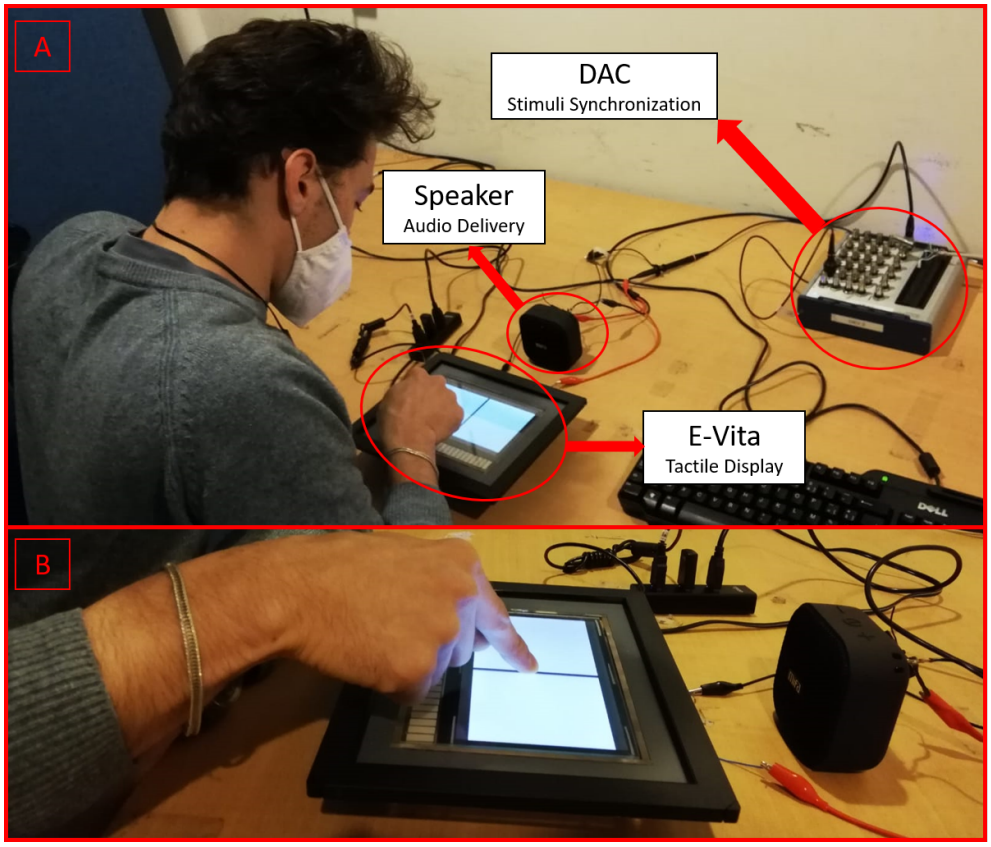

Figure 2.1: A) a participant during the experimental session with the Line condition. The main components (E-vita, speaker, and data acquisition card) are highlighted, and the experimental setup is illustrated. B) A zoom on the participant's finger performing the required task.

The selected signal for auditory feedback was a short burst of white noise with a duration of $50\ ms$. The white noise signal was selected to avoid performance biases due to the choice of a specific frequency. Indeed, a participant may be better than others on the SJ task due to a more sensitive hearing at specific frequencies rather than a higher ability for synchronicity discrimination.

In order to ensure a similar perceptual magnitude for the auditory and tactile stimuli, thus avoiding biases due to one stimulus being perceived much stronger than the other, the intensity of both stimuli was matched before the actual data collection started. To do so, we used a classical staircase with constant tactile and varying auditory intensity. Participants were asked to report which of the two stimuli (delivered simultaneously) was stronger. The staircase procedure ended after 10 reversals occurred and the auditory stimuli intensity was set to the average of the last 9 reversals for the remainder of the experiment. Finally, once triggered, the audio is delivered with or without a selected

delay, depending on the trial. More information on the selected delays is provided in the next section.

Visual condition

In order to investigate the influence of the presence of a visual cue on audio-tactile stimuli perceived synchronization, two visual conditions (V) were selected for our experiment:

- Black Screen (B): this first visual condition represents our control condition. In this case, no visual cue was provided to the participants, and the screen was a full-black picture.
- Line (L): this second visual condition was a white picture with a black line in the middle. This display was located just below the haptic surface to provide a visual cue that overlapped the tactile zone (see 3.3.B)

2.2.2.3 Experimental Protocol

Participants

We recruited 14 participants (six males and eight females, Age M=28.43 STD=6.72) for this experimental study. Ten participants were right-handed, and four were left-handed.

All participants confirmed they had no current history of neurological or psychiatric disorders and no loss of skin sensitivity in their hands. All participants confirmed not to have any hearing loss or disorders. The study received ethical approval from the Comité d'Ethique Hospitalo-Facultaire Saint-Luc UCLouvain, and all participants provided written informed consent and were remunerated for their participation in the study.

We excluded the data of one participant due to an issue with the velocity recording system.

Procedure

Participants sat comfortably in front of the setup in a quiet, dimly lit room. We then allowed participants to test the device to familiarize themselves with the system. They explored the tactile plate using the index of their dominant hand. To investigate whether the direction of the movement led to a bias in the perception of synchronicity, participants alternated between left-to-right (LTR) and right-to-left (RTL) swipes. For every sliding gesture, a tactile stimulus was delivered in the middle of the tactile plate (independently of the sliding direction). When the participant's finger reached the active zone (middle of the tactile plate), an auditory stimulus was delivered with different SOAs,

chosen during preliminary tests (0, 25, 50, 75, 100, 150, 200, 250, 300, 350, 400, 500, 600, 700 ms). The SOA was varied on a trial-by-trial basis using the psi method, a Bayesian adaptive algorithm that optimizes stimulus placement to maximize information gain about the PF parameters [154].

Before performing the sliding task and starting the experiment, participants were asked to use talc powder on the index of their dominant hand. The talc powder was used as it allows an homogeneous friction reduction independent of the finger's velocity, as highlighted by the results of Weal et al.[160]. Participants performed an SJ task where, at the end of each slide, they had to state whether they perceived the tactile and auditory stimuli to be synchronous or not. Moreover, the order of the two visual conditions was counterbalanced across participants to avoid biases.

In total, the experimental session comprised four independent psi procedures, two without a visual cue (LTR_B and RTL_B) and two with a visual cue (LTR_L and RTL_L). Thus, a value of threshold and slope was estimated for each movement direction and visual condition separately.

2.2.2.4 Measures and Analyses

Model choice

Since, within participants, separate psi algorithms were run for the four experimental conditions. The resulting estimates of threshold and slope parameter values obtained with this method assume independence between conditions. In particular, the same participant performing two different conditions is treated as two different participants. This assumption does not seem reasonable, as participants can be expected to have a baseline threshold/slope that is moderately modulated by conditions rather than a threshold/slope that is completely driven by each condition. As a result, it is difficult to distinguish the parts of the estimates obtained with the psi method associated with the participants and those associated with the conditions. To address this issue, we fitted a new model that takes the entire structure of the data into account (conditions within participants, participants within the population). Hierarchical models are well-suited for such structured data [161]. In this case, we used a hierarchical multiple probit (i.e., cumulative normal) regression model[162].

The threshold of the PF was modeled as a linear combination of subjects' random intercepts and fixed and random slopes for the factors *visual condition* (V), *movement direction* (D) and the interaction *visual condition* x *movement direction* (VxD). The slope of the PF was modeled as the base 10 exponential of a similar linear combination (constraining it to be positive). The guess and lapse

rates were treated as nuisance parameters, varying between participants but not between conditions.

Model fitting and diagnosis

A total of 438 stimulus-response pairs, coming from 13 participants, and four conditions were used to fit our model with the Rstan toolbox [155] [163]. Five chains of 4000 iterations (including a warm-up period of 2000 iterations, which was discarded) were used, leading to a total of 10000 draws from the posterior distribution. Generic non-informative gaussian hyper-priors were used for the threshold ($\mu_\alpha = 0$ and $\sigma_\alpha = 100$, except for the threshold intercept for which $\sigma_{\alpha_{LTR_B}} = 200$) and for $\log_{10}$(slope) ($\mu_\beta = 0$ and $\sigma_\beta = 0.5$) coefficients. Priors for the guess and lapse rates were beta distributions selected to keep most of the probability mass close to 0 but allow for non-zero values ($\alpha = 1$, $\beta = 50$).

Appropriate sampling of the posterior distributions was assessed using the ShinyStan (Stan Development Team, 2017) package based on the following diagnostic criteria: absence of divergent transitions; no reaching of maximum tree depth; good alignment of energy diagnostic plots, E-BFMI larger than 0.2; adequate sample size larger than 10% of the total sample size; Monte Carlo standard error smaller than 10% of the posterior standard deviation.Individual participants' data-fits were inspected visually as a posterior predictive check (Fig. 2.2).

Between conditions differences

The presence of significant effects of factors V, D or VxD was determined using bootstrapping of the posterior coefficient estimates (10^4 samples per test, two-sided tests). Pairwise comparisons were conducted using the same bootstrapping approach on posterior marginal means estimates. Lastly, to better illustrate the PF fitted from the posterior distribution of the model parameters, an expected PF was constructed for each condition. This was achieved by constructing 10^4 PFs using random draws from the posterior distribution of the parameters and taking, for each temporal delay, the 50^{th} percentile of the stimulus detection probability. To get a sense of the uncertainty around these values, the 2.5^{th} and 97.5^{th} percentiles (95% highest probability density intervals) are also plotted. These PFs can be interpreted as the expected values for a new unobserved participant coming from the same population.

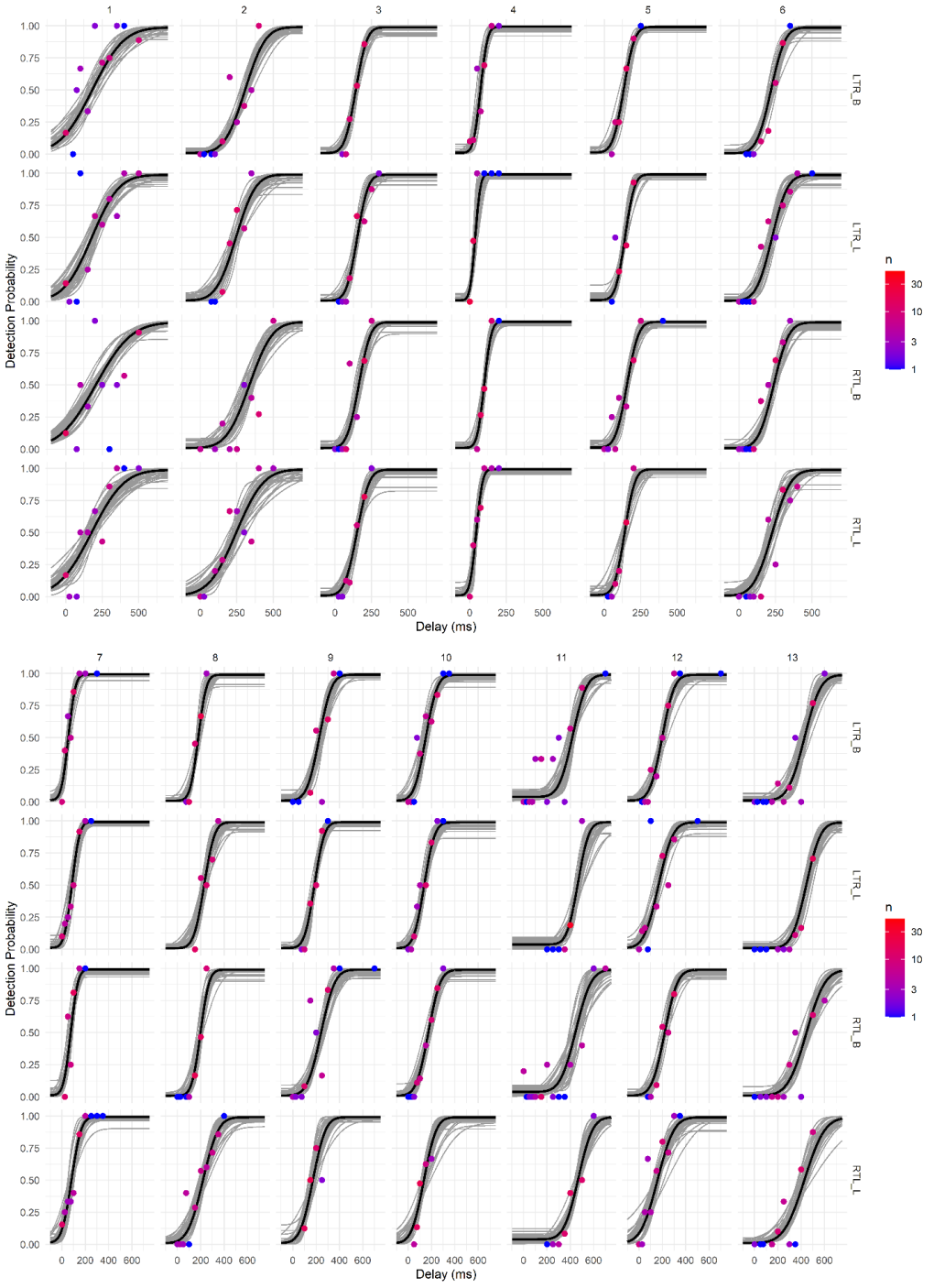

Figure 2.2: Individual participant fit of the model. For each participant (columns) and condition (rows), the detection probability for the various delays is depicted as dots, color-coded based on the number of trials used to compute the plotted probability. The PF constructed using the most likely parameter (median) is plotted in black. 100 PFs constructed using random draws from the posterior distribution of the parameters are plotted in grey (uncertainty around the true parameter values).

Two hierarchical models were fitted to the data, one using the temporal delays (ms) between tactile and auditory stimuli and one using the distance (mm) between the center of the screen (tactile stimulus location) and the position of the finger when the auditory stimulus was delivered.

A repeated measure ANOVA was used to assess whether the finger velocity differed between conditions. We also perform a statistical analysis (repeated measure ANOVA) on the normalized threshold (slope) to see if the finger velocity has an influence on the estimated parameters. The normalization was performed to have threshold (slope) values in the range from 0 to 1 and, therefore, compare delays with distances.

2.2.3 Results

The model diagnostics revealed appropriate sampling, and the model appeared to fit well with the participants' raw data. We show all the participants result in Fig. 2.2.

Threshold and slope

Posterior probabilities of effects of *visual condition* and *movement direction* on threshold revealed a main effect of *movement direction* ($p = 0.003$), as well a significant *visual condition* x *movement direction* interaction ($p = 0.014$). No main effect of the *visual condition* was identified.

During pairwise comparison, the thresholds for the black screen conditions with movements from left-to-right or right-to-left appeared to be significantly different ($p = 0.003$). All other comparisons were not significant.

A summary of posterior threshold parameter values for the different conditions is shown in Table 2.1.

Table 2.1: Posterior estimates for group-level threshold in the different experimental conditions. Where B (Black) and L (Line) represent the visual conditions and LTR (left-to-right) and RTL (right-to-left) represent the movement direction.

Visual Condition	Movement direction	Threshold (ms)	[95% HPDI] (ms)
B	LTR	211.73	[140.08; 283.65]
L	LTR	210.26	[136.48; 296.71]
B	RTL	233.30	[173.13; 315.60]
L	RTL	207.23	[141.80; 294.43]

Posterior probabilities of effects of *visual condition* and *movement direction* on slope revealed no main effect of *movement direction*, no main effect of *visual condition*, and no *visual condition* x *movement direction* interaction.

A summary of posterior slope parameter values for the different conditions is shown in Table 2.2.

Table 2.2: Posterior estimates for group-level slope in the different experimental conditions. Where B (Black) and L (Line) represent the visual conditions and LTR (left-to-right) and RTL (right-to-left) represent the movement direction.

Visual Condition	Movement direction	Slope	[95% HPDI]
B	LTR	0.014	[0.010; 0.020]
L	LTR	0.015	[0.010; 0.022]
B	RTL	0.013	[0.009; 0.019]
L	RTL	0.012	[0.008; 0.018]

We illustrate in Fig. 2.3 a visual representation of the model fitting to the expected PF from a new participant from the same population, with both median values and uncertainties.

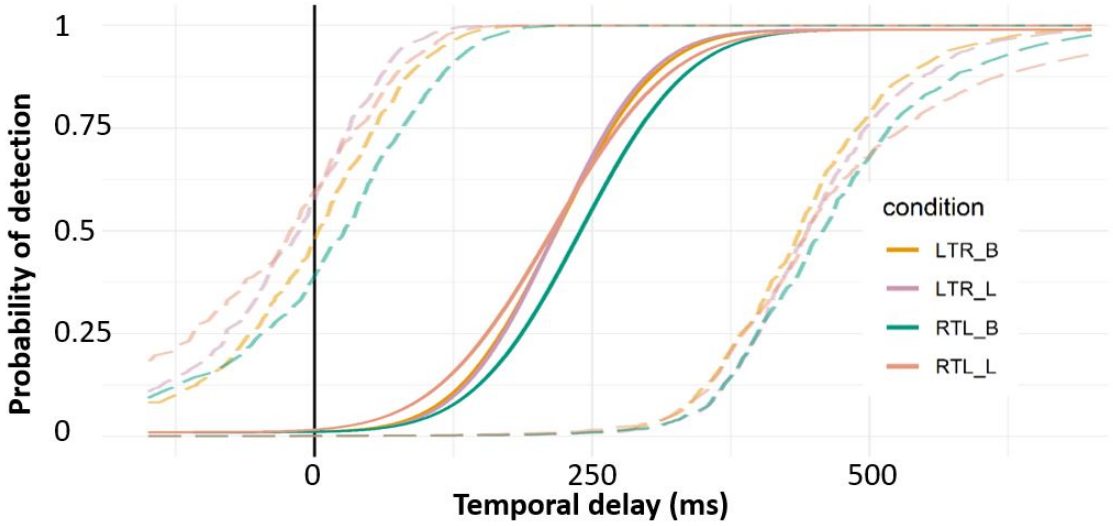

Figure 2.3: Posterior expected psychometric function. Bold lines represent the PF constructed with the most likely parameter values (i.e. the median of 10000 random draws from the posterior distribution of the model parameters). Dotted lines represent the uncertainty around the posterior expected PF (95% highest probability density intervals). The estimated threshold for our conditions is at 50% (i.e. 0.5 in the vertical axes) of probability of detection.

Finger Velocity

Overall, participants' finger velocities appeared independent of the *visual condition* or *movement direction* with a relatively small standard deviation, as depicted in Table 2.3.

By performing a repeated measure one-way ANOVA, our analysis did not reveal any statistical difference for finger velocity among the four conditions ($p = 0.0567$, $p = 0.5184$, $p = 0.1920$).

Table 2.3: Finger velocities: Mean and standard deviation for all our conditions

	LTR_B	RTL_B	LTR_L	RTL_L
Mean (mm/s)	27.9148	29.7789	28.5927	29.4776
STD (mm/s)	2.4440	2.3008	3.0481	3.2110

This result suggests that participants used a finger velocity that was not dependent on the condition. Therefore, the perceived delays are not influenced by the exploration velocity of participants but depend only on the temporal delay values between the two stimuli.

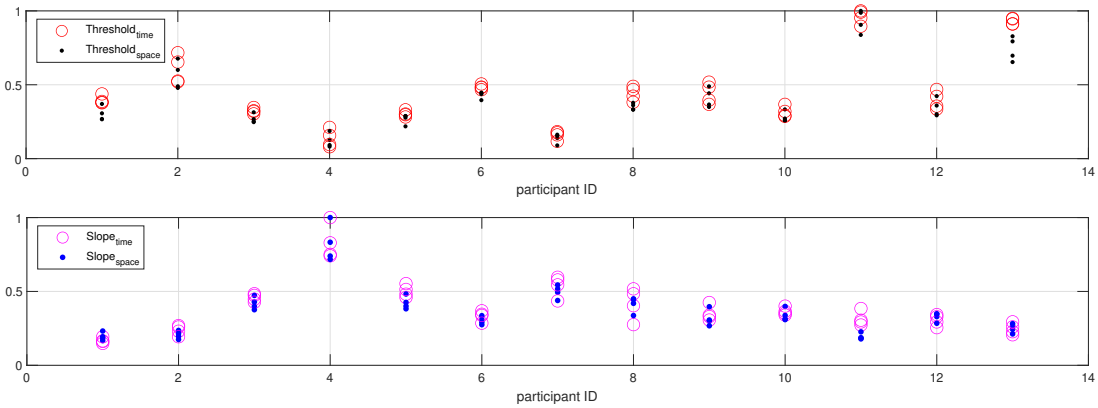

Figure 2.4: Normalized threshold (top) and slope (bottom) values for each participant and condition. This comparison is between values calculated considering time delay or space (i.e., time delay and sliding velocity).

To confirm the independence of the sliding velocity, we compare the normalized values of threshold and slope estimated with both delay (ms) and distance (mm), as shown in Fig. 2.4.

A visual inspection and a statistical analysis (repeated measure one-way ANOVA) on the normalized threshold and slope calculated in time and space confirmed the independence of the finger velocity. Hence, analyzing the performance of participants in relation to the spatial distance between stimuli does not provide additional information compared to the temporal delay analysis.

2.2.4 Discussion

Using psychophysics, we explored audio-tactile temporal interactions under conditions of active touch. To our knowledge, this is the first experiment investigating the maximum delay necessary for auditory and tactile stimuli (delivered during free exploration of a tactile display) to be perceived as synchronous in conditions of active touch.

Our results show that movement direction influences the ability of participants to perceive these delays, with right-to-left swipes leading to significantly larger thresholds than left-to-right swipes. This difference could be related to the fact that participants reported swiping their finger in that direction to be easier, possibly because, for all participants, left-to-right was the direction of writing. To dig deeper, it would be interesting to compare these results with an experiment where the finger slides in the proximal-distal direction. However, this change would introduce the problem that sliding

in one direction would cover the visual stimuli (as the hand would be over the screen), contrary to the opposite sliding direction.

Interestingly, the presence of a visual cue, indexing the position of the tactile stimulus, did not seem to systematically alter the ability of participants to detect temporal delays between the stimuli. However, the effect of the direction of movement seem to be mediated by the presence or absence of the visual cue, the difference between RTL and LTR conditions being much more pronounced in the black screen condition than in the line condition. This could indicate that in familiar situations (LTR), the individuals don't use the additional visual information but that they do in situations for which they have less prior experience or skills.

The acceptable delay we found in our study ($200\ ms$) is much greater than what was found, for example, in [147] ($20\ ms$) with a factor of 10. This large difference could be partially but not exclusively explained by differences between the tasks used by previous authors and ours. Moreover, another important difference is that the tactile condition we investigated is in active touch, which may involve processes that are not activated when passive touch is under investigation. Moungou et al. [164], as an example, by comparing brain activation using EEG responses, demonstrated that in passive touch the EEG response magnitude, elicited by a 22 Hz tactile signal, was significantly higher during passive touch compared to active touch. These indicate that the active exploration of the fingertip had an effect on the input-output transform resulting from the neural processing of that somatosensory input, considering that the somatosensory input generated in both active and passive touch was closely matched. Moreover, it has been shown that voluntary movement could affect temporal perception. Indeed, movements could result in a bias or a more precise temporal estimation, as reported by [165]. Nevertheless, the processes that govern how we perceive time across different senses still need to be better understood [153] and require further investigation.

Regarding the effect of finger sliding velocity, our empirical data show that participants had similar velocities regardless of the visual condition or movement direction. This result suggests that the finger exploration speed, chosen freely by the participants in our active touch experiment, may be related to the SJ task rather than the movement direction or the visual condition. We believe that participants chose the speed that allowed them to obtain the maximum information related to audio-tactile synchronization independently of the condition. Similarly, in rough macro-textures, the finger

velocity does not affect perception [166]. However, in our experiment, no textures were explored, and the tactile information was informative of the event itself (I felt something on the fingertip). Indeed, the finger sliding velocity's independence may be due to the nature of the tactile stimuli we chose. Moreover, by translating time delays into spatial distances (mm) and considering sliding velocity at each trial, pairwise comparisons did not reveal statistically significant differences for both visual conditions and direction of movement (for both slope and threshold), confirming the idea of finger speed independence. Nevertheless, our result is important as it shows a dissociation between the finger speed and the audio-tactile synchronization.

After this first experiment on audio-tactile synchronization while sliding the finger over an actuated surface, our interest will move onto audio-tactile synchronization with a virtual button. Moreover, we will dig even more by analyzing the data coming from the same type of task while comparing blind and sighted participants.

2.3 Temporal detection threshold of audio-tactile delays with virtual button

The following section is drawn from the journal paper of Brahimaj et al. [143].

Synchronization of audio-tactile stimuli represents a key feature of multisensory interactions. However, information on stimuli synchronization remains scarce, especially with virtual buttons. This work used a click sensation produced with traveling waves and auditory stimulus (a bip-like sound) related to a virtual click for a psychological experiment. Participants accomplish a click gesture and judge if the two stimuli were synchronous or asynchronous. Delay injection was performed on the audio (haptic first) or the click (audio first). In both sessions, one stimulus follows the other with a delay ranging from $0 - 700\ ms$. We used a weighted and transformed 3-up/1-down staircase procedures to estimate people's sensitivity. We found an asynchrony threshold of 179 ms and 451 ms for the auditory first and haptic first conditions, respectively. Statistical analysis revealed a significant effect between the two stimuli' order for threshold. Participants' acceptable asynchrony decreased when the delay was on the haptic rather than on the audio. This effect could be due to the natural experience in which the stimuli tend to be first tactile and then sonorous rather than the other way around. Our

findings will help designers create multimodal virtual buttons by managing audio-tactile temporal synchronization.

2.3.1 Related Work

Traditional push-button user interfaces are now replaced by touchscreens in most commercial devices. This solution offers versatile and customizable ways to design Human Machine Interfaces. However, it requires a high cognitive load, as it relies essentially on visual feedback. To cope with this issue, haptic surfaces [46] offer the possibility to rely less on sight by using the sense of touch in the communication process with the machine.

For example, vibrotactile feedback, which consists of the vibration of the touched surface, has been proven to be effective for typing on a smartphone or a tablet, increasing the performance and reducing typing errors [167][168]. However, more is needed to address all the perceptual aspects involved when pressing a physical button, such as the force feedback produced on the fingertip.

Different technologies have then been developed to overcome this limitation and, therefore, do not rely only on pure vibrotactile feedback. Tashiro et al. [169] proposed modulating the friction between an ultrasonically vibrating surface and the finger to recreate the sensation of rapid force changes typical of physical buttons. Similarly, Monnoyer et al. [170] modulated friction via ultrasonic vibration to induce a tactile sensation similar to a keystroke. The authors find that a mechanical detent in the case of a high level of friction followed by a low frictional level was perceived unambiguously by the participants. They also reported a weaker effect when the frictional levels were inversed, suggesting that the keyclick sensation was due to a release of skin stretch stored during the high friction state. This suggests that skin stretch due to lateral forces may play an important role in click perception. Following this idea, more recently, Garcia et al. [171] proposed a method that superimposes two vibration modes simultaneously (a longitudinal and a transverse mode) on a plate elicited by piezoceramic actuators in order to recreate a keyclick sensation. This superimposition creates an elliptical motion of the plate's particles, which is able to induce a lateral force on the fingertip. The keyclick sensation is obtained thanks to the elliptical motion inversion. The authors also performed a psychophysical study showing that the keyclick haptic feedback was perceived by all their participants [172].

Progressive ultrasonic waves to generate keyclick sensation on an actuated surface have also been proposed [173]. This method consists of a traveling wave based on a predefined force threshold that is reversed when a second force threshold is reached. With this method, the frictional forces produced by the traveling waves can deliver the keyclick sensation when the actuated surface is pressed. The authors performed an interesting comparison of their method with Tashiro's modulation of friction. The results show that both methods are similar in terms of the quality of the sensation. However, Gueorguiev's method is promising for all those applications where a press is performed and users do not perform any lateral movement of the finger. Therefore, we decided to use their method for our haptic feedback.

Even if new methods now exist for push-button generation that do not rely purely on vibrotactile feedback, the haptic modality alone, however, can not resemble the sensation of a physical button. A physical button is intrinsically bi-modal considering the haptic feedback produced by the pressing gesture and the associated click sound it emits, and it has been shown that both haptic and auditory contribute to the perceived tactile strength [174]. Therefore, it is important to employ this bi-modal relationship in virtual buttons to improve how they are perceived by the user.

Investigating this bi-modal relationship, Kaneko et al. [175] proposed a pseudo-haptic method to vary the sensation of heaviness while participants were performing a click by modulating the auditory feedback. Their result suggests that by presenting a low-frequency pure tone in response to the user clicking a button, they were able to increase the heaviness sensation. This pseudo-haptic effect is simple to implement but still, it can improve the way we perceive a virtual button.

Among different factors related to the integration of multimodal information, temporal aspects represent a key element for perception. Indeed, the processing time for the modalities can be different, and the perception can be affected if they are not precisely synchronized. As the asynchronies between two modalities increase, the sense of realism or the user comfort decrease [176], while simultaneous stimuli give rise to better performances as reported by [104]. The latency between modalities can limit the quality, effectiveness, and interactivity of virtual buttons.

To shed light on this interaction, many works have explored the perceived simultaneity of audio-tactile stimuli in different contexts. Fujisaki et al. [148] investigated temporal resolution of synchronous perception for different modality pairings where stimuli were a light blob, a white noise,

and a vibration on the index finger. The authors reported the audio-tactile superiority in temporal resolution over visio-tactile and audio-visual combinations. Occelli et al. [153] have reviewed prior works describing the results of behavioral studies investigating temporal resolution between hearing and touch. The author highlighted the increased interest in the field and the need for more investigation into audio-tactile temporal perception. Hence, understanding the temporal perceptual aspects of audio-tactile interactions is important when designing multimodal experiences.

In this context, Hao et al. [177] investigated the simultaneous perception of audio-tactile stimuli during voluntary movement with a temporal order paradigm. For this study, participants were moving their index fingers from left to right and were not in the condition of touching a virtual button or performing an exploration with their finger. The results show that voluntary movement affects the point of subjective simultaneity (PSS) but does not influence the just noticeable difference (JND). This result is important as we are interested in the temporal aspect of audio-tactile interactions and voluntary movement is intrinsically related to the action of pushing a button.

While various research has investigated the perceptual tactile aspect of virtual buttons, less has been done concerning the auditory stimulus and virtual buttons. A relevant work on virtual buttons was performed by Kaaresoja et al. [178]. The authors have investigated touch simultaneity with audio, tactile or visual feedback. During their experiment, the authors varied the latency of the feedback (tactile, audio or visual) with respect to the momentthe participant was touching a virtual button simulator but not receiving any type of tactile feedback. The authors found that the latency should be lower than $50\ ms$ for tactile, $70\ ms$ for audio, and $85\ ms$ for visual feedback. Even if relevant, this study is in unimodal condition as only one feedback at a time was presented and, therefore, does not give insight into the temporal simultaneity of the audio-tactile interaction that is of our interest. It is, therefore, essential to conduct this type of study in the specific case of virtual button clicks with auditory feedback. Indeed, to our best knowledge, no research has been conducted on temporal aspects related to virtual buttons in the audio-tactile bi-modal condition.

2.3.2 Method

Two experiments were conducted to measure the point of subjective simultaneity of auditory-tactile and to estimate the psychometric function. In the first experiment (HF), the haptic stimulus was

always presented before the audio stimulus, while in the second one (AF), audio was always delivered before the haptic stimulus. In both experiments, the second stimulus followed the first one with a stimulus-onset asynchrony (SOA) ranging from 0 to 1000 ms with a fixed step size of 15 ms.

Each participant underwent both experiments, and we alternated the order to avoid bias due to the presentation order.

A weighted and transformed 3-up/1-down staircase was employed to present the delay at each trial. The calculation of the targeting level can be obtained by using the following formulas:

$$P_{down} = p^N, \quad P_{up} = 1 - p^N \tag{2.1}$$

where p_{up} and p_{down} are the probabilities of responses leading to the staircase going up or down, p is the target fraction correct level, and N is the number of correct responses necessary to go down. In our case, N is set to 2. Considering the equilibrium condition for convergence

$$S_{down}\mathring{\mathrm{u}}P_{down} = S_{up}\mathring{\mathrm{u}}P_{up} \tag{2.2}$$

where S_{up} and S_{down} are the up and down step sizes, and by substituting (1) in (2), we obtain:

$$S_{up}/S_{down} = p^N/(1 - p^N) \tag{2.3}$$

With a 3-up/1-down algorithm, the upward step size is three times bigger than the downward step size, leading to $S_{up}/S_{down} = 3$, targeting $p = 0.866$. Hence, our weighted and transformed 3-up/1-down staircase targets the 86.6% perceptual threshold [179].

2.3.2.1 Experimental Setup

Our system (see Fig. 2.5) comprises two microcontrollers (STM32F4) handling one stimulus each. The microcontroller (μC) dedicated to the tactile stimuli was in charge of producing both the traveling wave and the change of direction of the wave. Moreover, the audio stimulus was triggered when the threshold force was reached (haptic first - HF). In the case of audio first (AF), an internal timer was set to provide the desired delay on the keyclick after the audio stimuli.

The second μC was in charge of audio delivery. In the case of AF, the short burst was delivered

as soon as the μC was triggered. In contrast, in the HF case, an internal timer was responsible for providing the audio feedback with the desired delay after the tactile stimuli.

In order to allow a good synchronization between the stimuli, we performed a different test on both HF and AF conditions. The measurements revealed an internal mean latency of $100\,\mu s$ over the multiple tests. Such value is acceptable for the type of experiment, as it allows a low latency synchronization between the stimuli.

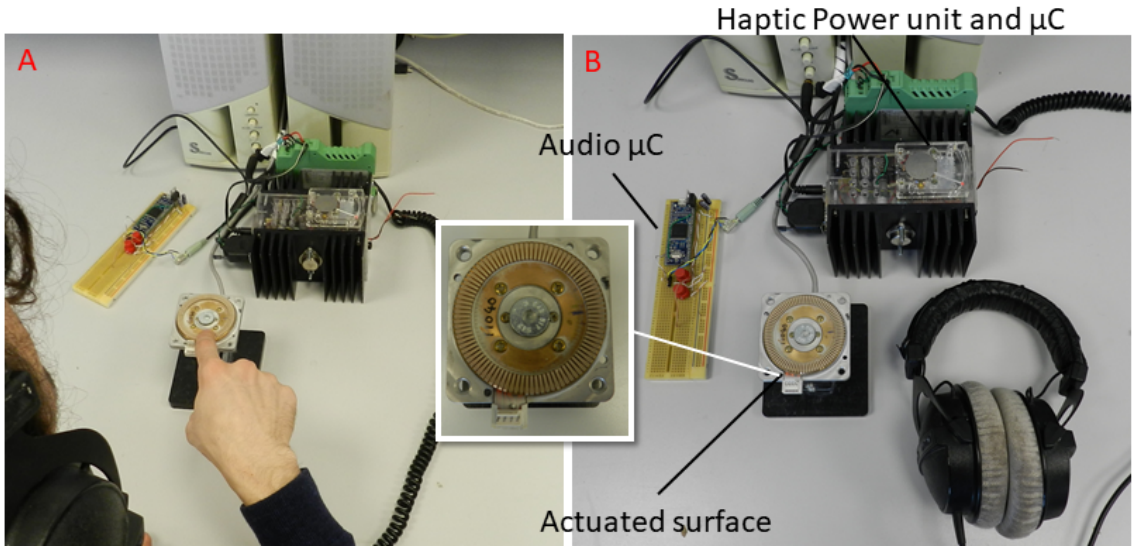

Figure 2.5: A) Participant during the experiment. B) System composed of the actuated surface (USR60), the haptic power unit with the microcontroller, and the microcontroller used for the auditory feedback.

2.3.2.2 Stimuli

Tactile

The tactile stimuli were delivered by a haptic device developed by Gueorguiev et al. [173]. The device is based on the stator of an ultrasonic motor (USR60, Shinsei Corporation, Japan). It consists of a bronze beryllium disc under which a ring of 16 piezoelectric actuators is glued. Half of these actuators are arranged to excite a transverse mode, with a resonance frequency of 40 kHz and a wavelength of $\lambda = 21\,mm$. The remaining actuators, positioned with a spatial shift of $\lambda/4$, excite a doublet of the previous mode. The excitation of both vibration modes, with a temporal phase shift of $\pi/2$, generates a traveling wave. Changing the sign of the phase shift inverts the direction of the traveling wave, while the cancellation of the phase shift results in a stationary wave. To generate

the keyclick sensation using our haptic device, we modulate the traveling wave depending on a force sensor applied in the USR60. Indeed, when the produced finger-pushing force reaches a predefined threshold of $f_{\text{th}} = 0.25\ N$, we produce the traveling wave in a direction (Fig. 2.6.B1). Then, after a predefined time-delay $\tau = 10\ ms$, we reverse the direction of the wave (Fig. 2.6.B2). Finally, the change of direction of the traveling wave gives the keyclick sensation by a lateral force produced on the fingertip [173]. From the moment the force threshold was reached to the moment the reversed traveling wave was stopped, our haptic signal duration is $40\ ms$.

Auditory

Auditory stimuli are delivered to participants through a pair of headphones. The selected stimulus is a pure tone with 500 Hz frequency and 100 ms duration (see Fig. 2.6). For the volume level, we selected 30% in loudness concerning its maximum. We choose a low-frequency pure tone because, as suggested by [175], this can increase the perception of heaviness, making the click sensation more similar to classical buttons. The duration and intensity of the signal was chosen based on pretrial tests. In addition to the pure tone, a white noise sound is played constantly in the background (low intensity/low volume) to prevent participants from hearing any audible noise produced by the tactile device. The white noise was chosen because we wanted to compare the auditory stimuli with the tactile stimuli without the aid of the audible noise coming from the actuated surface.

2.3.2.3 Experiment Protocol

Participants

We recruited a total of seventeen participants. We excluded two participants due to exclusion criteria (numbness/loss of touch and loss of auditory). Fifteen healthy volunteers participated in these experiments (9 males and 6 females, aged $M = 29,2$ and $SD = 4.00$). 11 participants were right-handed, and four were left-handed. None of the remaining participants exhibited any motor difficulty or loss of sensation in their dominant hand index. Each participant took part in the experiments voluntarily and signed the informed consent before the start of the experiment, and the ethical committee of Lille University approved the experiment.

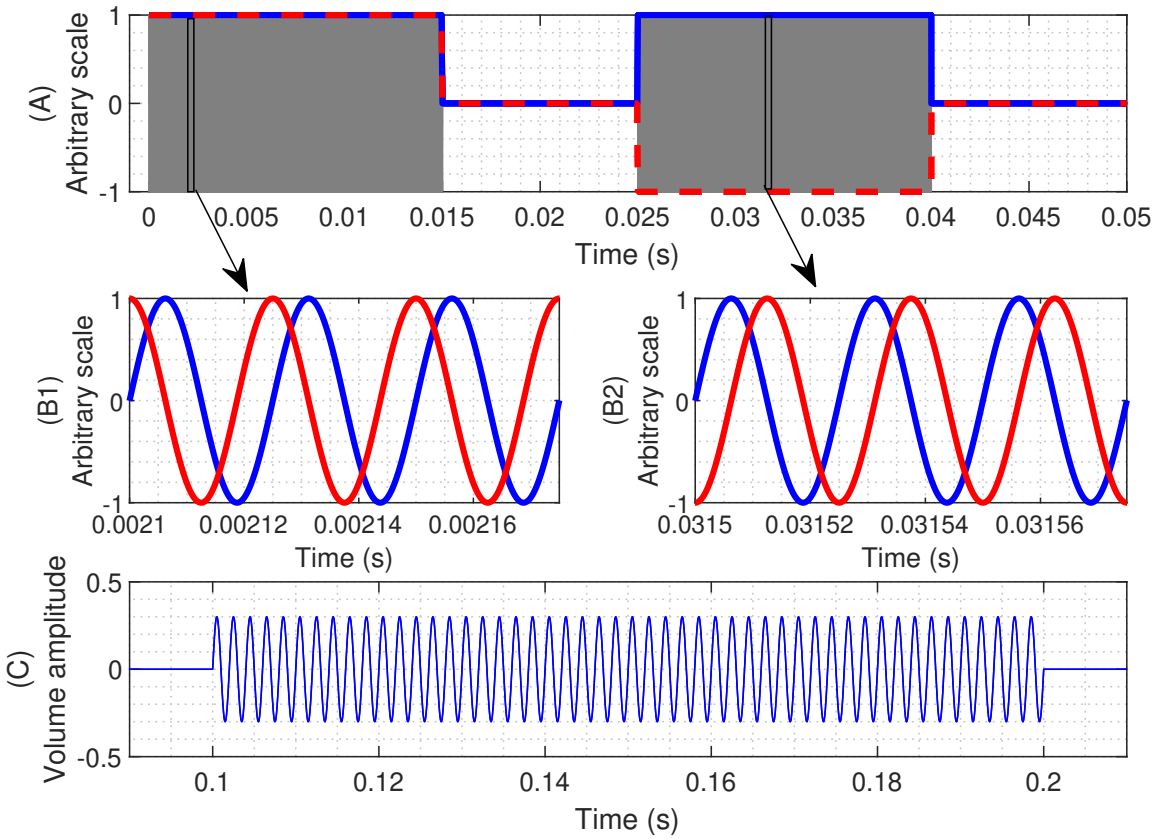

Figure 2.6: **A**) Te haptic signal used for the keyclick generation. It is composed of the superposition of two stationary waves depicted with their amplitude reference in blue and red colors. Once the force threshold is reached (time $= 0$), the traveling wave is switched on (0 to $15\ ms$) by a step increase in the amplitude vibration and the two waves are shifted by $+90°$ as depicted in B1. After $10\ ms$ the direction of the traveling wave is reversed, performing a $-90°$ phase shift ($25\ ms$ to $40\ ms$) as depicted in B2. **B1**,**B2**) shows a zoom of the two ultrasonic waves and their phase shift. **C**) Recorded waveform and timing of the audio signal used as auditory feedback in the condition were auditory was played with 100 ms delay compared to the tactile feedback.

Procedure

In both experiments, participants sat comfortably in a chair in front of the device, and the experiment took place in a quiet room. After reading and signing the informed consent form, we allowed participants to try the device and familiarize themselves with the haptic feedback by performing some keyclick. Before starting the experiment (either the first or the second), we presented to participants the stimuli with different SOAs: 0, 100, 200, 300, 400, 500, 600, 700, 800, 900, and 1000 ms. This

step was performed to ensure a good understanding of the experiment by participants.

Participants then performed a forced-choice task where, at each trial, they were asked to accomplish a click gesture with the index finger of their dominant hand. Each keyclick was performed only once. After pressing the actuated surface, they had to report whether they perceived the tactile and auditory stimuli to be synchronous or not synchronous before moving to the next trial.

All the participants performed a total of 100 trials for each experiment. The minimum number of reversal points we achieved with this number of trials during the experiments were 21 (M=30.3, STD=4.68) and 26 (M=32.2, STD=4.23) for HF and AF, respectively. This number ensures a good estimation of the parameters as we run a pretrial test in order to choose the initial value of the staircase to be close to the expected estimated threshold [180]. We chose random values between $300\ ms$ and $400\ ms$ for the HF condition and random values between $100\ ms$ and $250\ ms$ for the AF condition.

2.3.2.4 Data Analysis

The psychophysical curve was estimated with the weighted and transformed 3-up/1-down staircase. We estimated the threshold and slope of each participant in both experiments separately.

A standard function to predict the psychometric function for a forced-choice experiment is the so-called Weibull cumulative distribution function, defined as follows:

$$W(x) = 1 - (1-g)e^{-\frac{k.x}{t}^s}, with\ \ k = -\log\left(\frac{1-a}{1-g}\right)^{\frac{1}{s}} \tag{2.4}$$

where g is the guess rate, i.e., the expected performance to be achieved by chance (0.05 in our case because of the experimental design) which represents the lower asymptote of the PF. t represents the threshold, s the slope of the psychometric function, and a is the performance level or targeted threshold performance (86.6%).

The density function predicts the probability that the subject will answer that the two stimuli were asynchronous. Considering a trial i, with a delay x_{i}, the probability for the subject to answer asynchronous (P_a) or synchronous (P_s) could be written as:

$$P_{\text{a}} = W(x_i)\ \ or\ \ P_{\text{s}} = 1 - W(x_i) \tag{2.5}$$

Where $W(x_i)$ represents the Weibull for a given delay x_i. We can then define the probability of observing the whole experiment by the log-likelihood function, an efficient method for estimating the parameters of the psychometric curve [180] (known as Fisher's maximum-likelihood procedure). The function is defined as:

$$h_i \log(W(x_i)) + (1 - h_i) \log(1 - W(x_i)) \tag{2.6}$$

Where h_i is the subject's answer (1 or 0 if asynchronous or synchronous).

A chi-square χ^2 goodness of fit test was also performed. This test can be used to determine whether our observed frequency distribution deviates significantly from the hypothesized Weibull distribution (2.4). The test confirmed the hypothesized distribution and didn't show any significant deviation. Model parameters (threshold t and slope s) were then estimated by maximizing the log-likelihood equation.

All the data analysis was performed in MATLAB 2022b with the appropriate toolboxes [181].

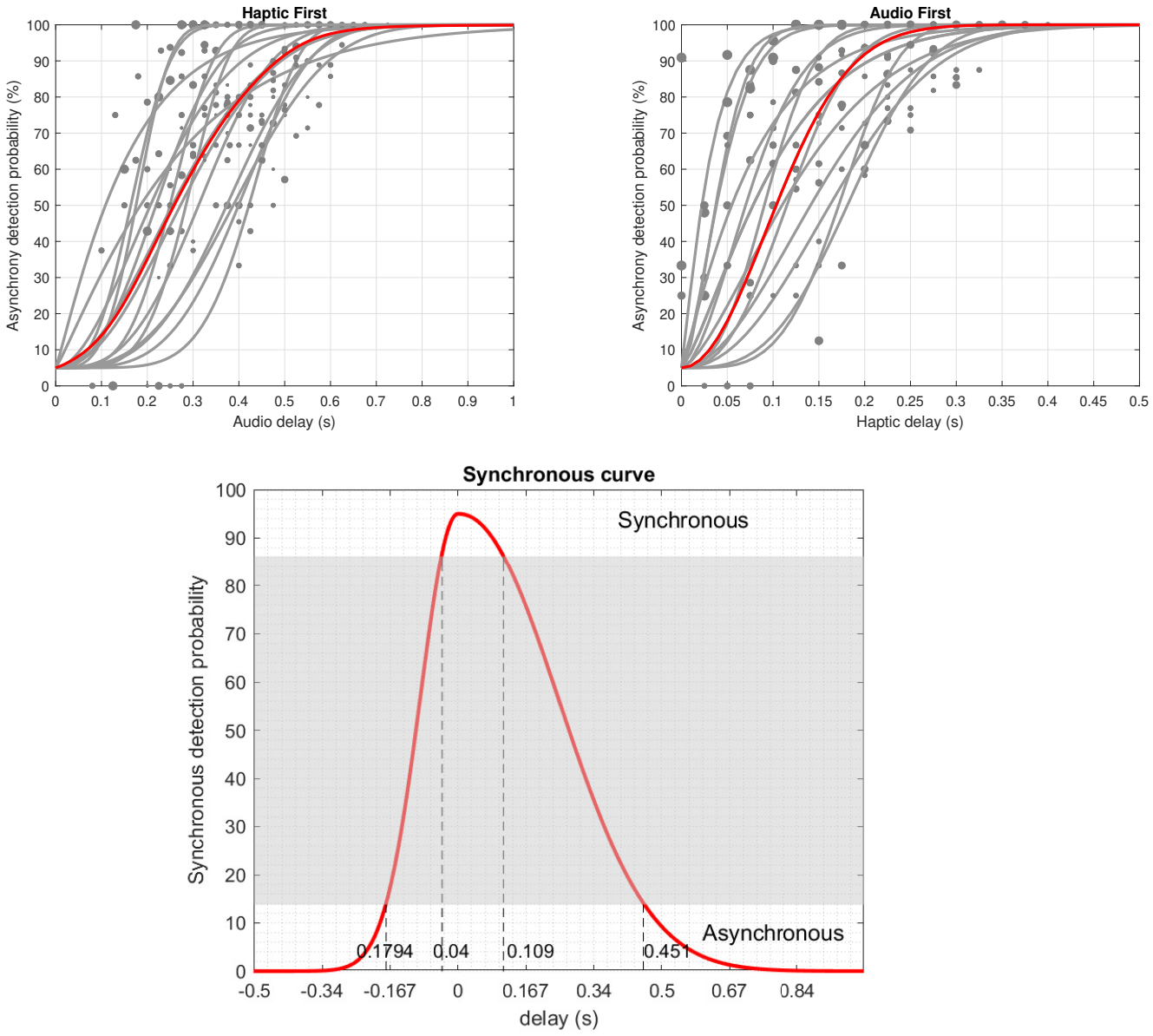

Figure 2.7: Up) Asynchronous discrimination performance in the haptic first and audio first condition. Individual estimated curves are plotted in light grey and data (grey dots) are plotted as the proportion of trials in which the two stimuli were judged as asynchronous as a function of the audio delay. The dimension of the dots is related to the occurrence of the associated delay. The psychometric curve constructed with the most likely parameter's value (medians) is represented in red and a higher value corresponds to a higher probability of asynchronous. **Down**) Synchronous curve obtained applying eq. 2.5. Negative delay values indicate AF condition and positive values indicate HF condition. High/low values indicate a high/low probability of feeling the stimuli as synchronous. The upper (lower) side of the curve, labeled as synchronous (asynchronous) represent delays with a high probability to be felt as synchronous (asynchronous).

2.3.3 Results

For each subject, we calculated the probability of the stimuli being felt as asynchronous as a function of the SOAs. We then estimated the best-fitting psychometric function for each individual by maximizing the log-likelihood function, as explained in the preview section. In Figure 2.8, we show the best-fitting threshold value for each participant and both HF (red) and AF (blue) conditions. After visual inspection, the PFs appeared to match well the collected data. The PFs are illustrated in Figure 2.7 by the curves plotted in light grey for the HF and AF conditions, respectively.

For the AF condition, the median value of threshold and slope calculated were respectively $179.4\,\text{ms}$ $(IQR = 0.231 - 0.134)$ and 1.755 $(IQR = 2.4353 - 1.3485)$. We used these values to construct the best-fitting psychometric curve depicted in Figure 2.7.

On the other hand, for the HF condition, the median value of threshold and slope calculated were respectively $451.0\,\text{ms}$ $(IQR = 0.50874 - 0.334)$ and 2.962 $(IQR = 3.8044 - 1.992)$. We illustrate the best-fit using these parameters in Figure 2.7.

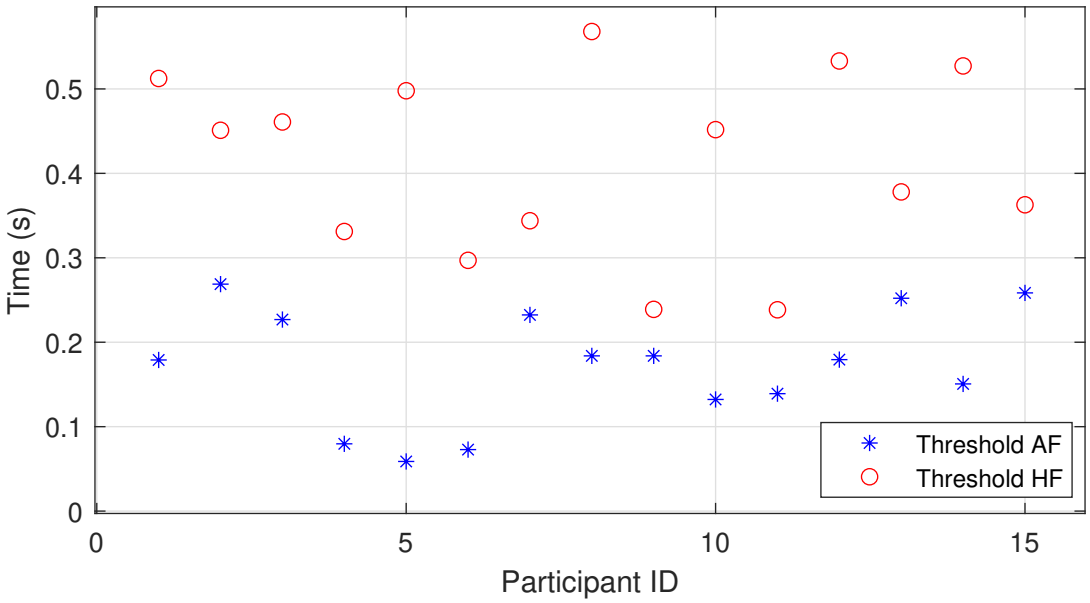

Figure 2.8: Best estimated threshold parameters for each participant and both haptic first (red circles) and audio first (blue stars) conditions.

The threshold values of $179.4\,\text{ms}$ and $451.0\,\text{ms}$ for the AF and HF condition, respectively, represent the delay that, with 86.6% probability, will make the audio-tactile stimuli be felt asynchronous. On

the contrary, the synchronous curve depicted in Fig. 2.7, obtained by applying equation 2.5, gives information on the simultaneity perception of audio-tactile stimuli. Here, the threshold values are 40 ms and 109 ms for AF and HF conditions, respectively. These threshold values represent the delay that will make the audio-tactile stimuli feel synchronous, with an 86.6% probability (i.e., with a 14.4% probability of the two stimuli feeling asynchronous).

Generally, threshold values in the AF condition tend to be lower than in the HF condition, as illustrated in Fig. 2.8. Furthermore, we compared the estimated thresholds in the HF condition with those estimated in the AF condition. By performing an ANOVA (Figure 2.9), we found that for both threshold and slope, there was a statistically significant difference ($p = 6.39 \times 10^{-8}$ for threshold and $p = 0.0167$ for slope) between the two conditions. This highlights an asymmetry between the conditions and the tendency of individuals to be more sensible to asynchrony in the condition where delays were injected into the haptic stimuli. Furthermore, our results also show a larger variance for both threshold and slope in HF condition with respect to AF.

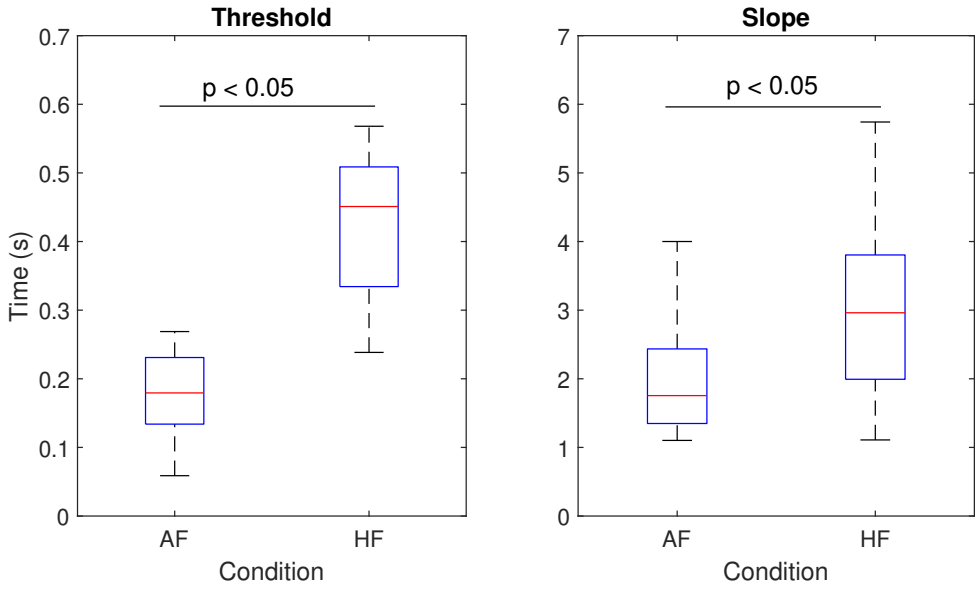

Figure 2.9: ANOVA box-plot for threshold and slope for both haptic first and audio first conditions

2.3.4 Discussion

In this study, we evaluated the temporal sensitivity of synchrony perception for audio-tactile stimuli with virtual buttons. We estimated threshold parameters where injection of delay was performed in the auditory stimuli preceded by the tactile stimuli and vice versa. We estimated the most likely parameter value (threshold and slope), and our results show that temporal sensitivity was different in the two conditions (HF, AF). Our analysis shows a statistically significant difference in threshold and slope between the two conditions.

The results show a tendency of individuals to be more sensitive to delay injection in the haptic modality rather than in the auditory. This difference may be due to our natural experience in our daily life. A multisensory experience is most likely to be felt first by our haptic modality and then followed by auditory. This sensation is true mainly when we perform a click gesture, where a button is first felt tactically and then heard (or followed by another audio event). We believe that this experience we live in our daily lives reinforces our expectation of one stimulus (haptic) followed by another (audio in our case). Therefore, it may create a sensation of discomfort or a negative surprise effect when the stimuli are inverted. Indeed, we think this mechanism is responsible for the higher sensitivity of asynchrony detection (i.e., lower threshold) in the audio first condition.

2.4 Audio-Tactile synchronization with a button click: comparison of Blind vs Sighted people

The experiment was conducted with the support of the IIT's laboratory members of the UVIP (Unit For Visual Impaired People). The primary aim of the experiment was to investigate how blind individuals perceive and synchronize auditory and haptic information when interacting with virtual buttons, and to determine the level of sensitivity to asynchronies.

This section meticulously outlines the research's groundwork, the applied methodology, and the results. Beyond that, it underscores the substantial implications these findings carry for the development of assistive technologies and tools, including touch screens, voice assistants, and other digital interfaces, designed to enhance the daily lives of individuals with visual impairments.

At its core, this experiment illuminates the intricate mechanisms underpinning sensory synchro-

nization, offering profound insights into how the integration of auditory and haptic information is influenced by the utilization of virtual buttons. As such, the outcomes of this research not only propel our comprehension of sensory perception but also fortify its applications within the realm of assistive technology, ultimately striving to make the world more accessible for individuals with visual impairments.

Figure 2.10: Device used for the experiment. On the right side we show the internal part of the case.

2.4.1 Method

One experiment was conducted to measure the point of subjective simultaneity of auditory-tactile and to estimate the psychometric function. In the first experimental condition (haptic first, i.e. HF), the haptic stimulus was always presented before the audio stimulus, while in the second one (audio first, i.e. AF), audio was always delivered before the haptic stimulus. In both conditions, the second stimulus followed the first one with a stimulus-onset asynchrony (SOA) ranging from 0 to 1000 ms with a fixed step size of 15 ms. Each participant underwent both experimental conditions, and we alternated the order to avoid bias due to the presentation's order.

2.4.1.1 Experimental setup and Stimuli

For the experiment, the device depicted in Fig. 2.10 has been developed. In the experiment, we did not use the device in section 2.3 as it was not available. However, this allows us to compare the result of sighted participants in this experiment with those in the previews section, taking in

consideration the different devices. The device depicted in Fig. 2.10 is composed of two independent microcontrollers, each of them responsible for the delivery of stimuli. In particular, a Nucleo-board (STM32F446 Nucleo-64) was used for the delivery of the auditory stimuli. The stimuli was a pure tone of 500 Hz, with a duration of 100 ms. The second microcontroller (BOS1901) was used for the control and actuation of a piezoelectric actuator (TDK Mini PowerHapTM 1204H018V060) that was providing the click sensation. The actuator, which has high acceleration and large forces in a very compact design and a short response time of less than one millisecond, has dimensions of 12 x 4 x 1.8 mm (illustrated in Fig 2.10).

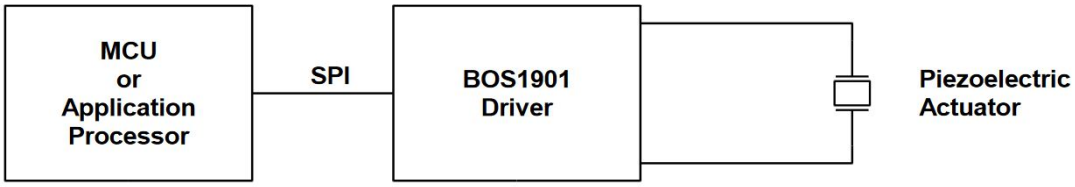

Figure 2.11: System components using piezoelectric actuator

The click actuation signal is depicted in Fig. 2.12. When no finger is pressing the component, the system is in SENSE mode. In this mode, the piezoelectric component is used as a force sensor, waiting for the force threshold to be reached. When this happens, the piezoelectric component is used as an actuator and the feedback signal waveform is played. When the signal is over, after a stabilization time, the actuator is again ready and set in SENSE mode.

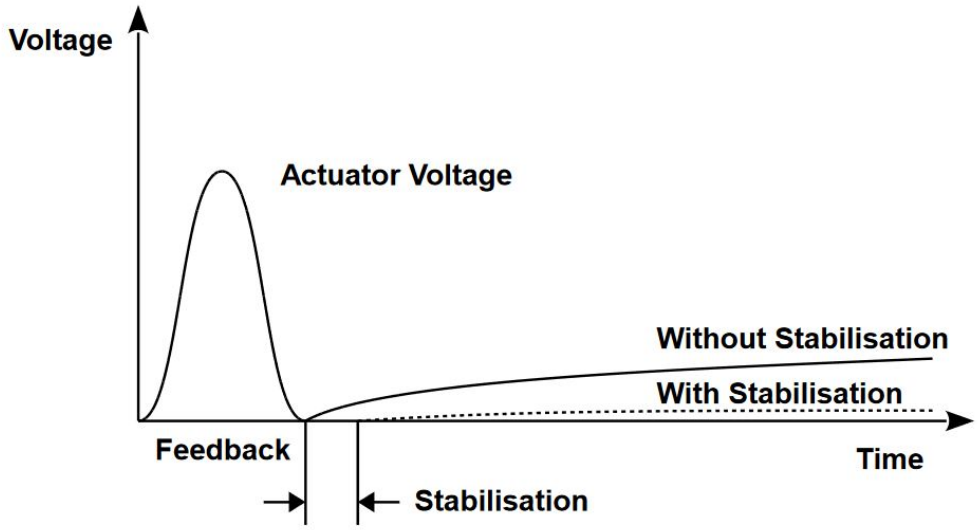

Figure 2.12: Piezo creep following feedback waveform

An important aspect of this experiment is related to the synchronization of the two stimuli from the device's point of view. Indeed, to ensure good synchronization, different latency measurements were performed comparing the signals on an oscilloscope. As a result, the two stimuli show a latency of about 800 μs to 900 μs. This latency is negligible for our experiment.

2.4.1.2 Experimental Protocol

Participants

We recruited a total of fifteen participants. We excluded two participants due to exclusion criteria (numbness/loss of touch and loss/poor of audition). Thirteen healthy volunteers participated in these experiments (7 males and 6 females, aged $M = 34,4$ and $SD = 4.52$). Seven participants were blind (4 females and 3 males) and 6 were sighted (3 males and 3 females). None of the remaining participants exhibited any motor difficulty or loss/lack of sensation in their dominant hand index or auditory system. Each participant took part in the experiments voluntarily and signed the informed consent before the start of the experiment, and the ethical committee of the Italian Institute of Technology approved the experiment.

Procedure

In both experimental conditions, participants sat comfortably in a chair in front of the device, and the experiment took place in a quiet room. After signing the informed consent form, we allowed participants to try the device and familiarize themselves with the haptic feedback by performing some keyclick. Before starting, we presented to participants the stimuli with different SOAs: 0, 100, 200, 300, 400, 500, 600, 700, 800,900, and 1000ms. This step was performed to ensure a good understanding of the experiment by participants. Participants then performed a forced-choice task where, at each trial, they were asked to accomplish a click gesture with the index finger of their dominant hand. Each keyclick was performed only once. After pressing the actuated surface, they had to report whether they perceived the tactile and auditory stimuli to be synchronous or not synchronous before moving to the next trial. All the participants performed a total of 100 trials for each experimental condition.

Data Analysis

The data collection was obtained by an algorithm that was a weighted and transformed 3-up/1-down staircase, targeting threshold performances at 86.6%. The estimation of threshold and slope parameters was performed for each participant in both experiments separately. Model parameters (threshold t and slope s) were then estimated by maximizing the log-likelihood function. All the data analysis was performed in MATLAB 2022b with the appropriate toolboxes and the algorithm for the analysis was written and developed at the University of Lille, at IRCICA. As the data analysis method and algorithm have been discussed and illustrated previously, we invite the reader to refer to section 2.3.2.4 for more details.

2.4.2 Results

Sighted participants

For each subject, we calculated the probability of the stimuli being felt as asynchronous as a function of the SOAs. We then estimated the best-fitting psychometric function for each individual by maximizing the log-likelihood function, as explained in the preview section. In Fig. 2.13, we show the best-fitting threshold value for each participant and both HF (red) and AF (red) conditions. After visual inspection, the PFs appeared to match well the collected data. The PFs are illustrated in Figure 2.13 by the curves plotted in light gray for the HF and AF conditions, respectively.

For the AF condition, the median value of threshold and slope calculated were respectively 235.8 ms and 2.012. We used these values to construct the best-fitting psychometric curve depicted in Figure 2.13. On the other hand, for the HF condition, the median value of threshold and slope calculated were respectively 397.6 ms and 2.392. We illustrate the best-fitting using these parameters in Figure 2.13.

The threshold values of 235.8 ms and 397.6 ms for the AF and HF condition, respectively, represent the delay that, with 86.6% probability, will make the audio-tactile stimuli be felt asynchronous. On the contrary, the synchronous curve depicted in Fig. 2.13, obtained by inverting the probability equation (i.e., $P_{synchronous} = 1 - P_{asynchronous}$), gives information on the simultaneity perception of audio-tactile stimuli. Here the threshold values are 42.3 ms and 145.5 ms for AF and HF conditions, respectively. These threshold values represent the delay that will make the audio-tactile stimuli feel synchronous, with an 86.6% probability (i.e., with a 14.4% probability of the two stimuli feeling asynchronous).

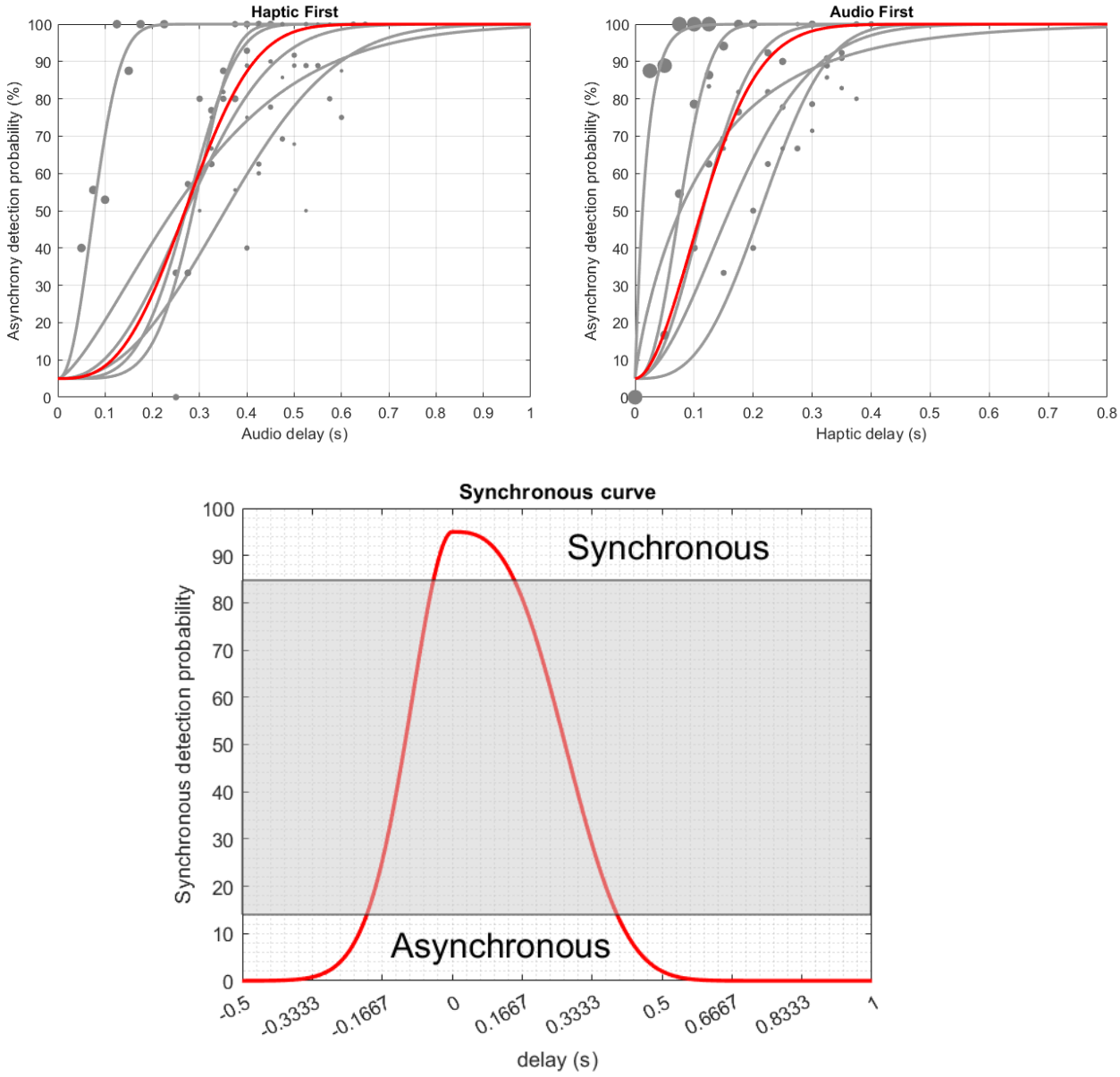

Figure 2.13: Up) Asynchronous discrimination performance in the haptic first and audio first condition for sighted participants. Individual estimated curves are plotted in light gray and data (gray dots) are plotted as the proportion of trials in which the two stimuli were judged as asynchronous as a function of the audio delay. The dimension of the dots is related to the occurrence of the associated delay. The psychometric curve constructed with the most likely parameter's value (medians) is represented in red and a higher value corresponds to a higher probability of asynchronous. **Down**) Synchronous curve obtained calculating the probability of a synchronous response. Negative delay values indicate AF condition and positive values indicate HF condition. High/low values indicate high/low probability to feel the stimuli as synchronous.

Blind participants

For each subject, we calculated the probability of the stimuli being felt as asynchronous as a function of the SOAs.We then estimated the best-fitting psychometric function for each individual by maximizing the log-likelihood function, as explained in the preview section. In Fig. 2.14, we show the best-fitting threshold value for each participant and both HF (red) and AF (red) conditions. After visual inspection, the PFs appeared to match well the collected data. The PFs are illustrated in Figure 2.14 by the curves plotted in light gray for the HF and AF conditions, respectively. For the AF condition, the median value of threshold and slope calculated were respectively 365.1 ms and 2.286.

We used these values to construct the best-fitting psychometric curve depicted in Figure 2.14. On the other hand, for the HF condition, the median value of threshold and slope calculated were respectively 596.9 ms and 2.315. We illustrate the best-fitting using these parameters in Figure 2.14.

The threshold values of 365.1 ms and 596.9 ms for the AF and HF condition, respectively, represent the delay that,with 86.6% probability, will make the audio-tactile stimuli be felt asynchronous. On the contrary, the synchronous curve depicted in Fig. 2.14, obtained by inverting the probability equation (i.e., $P_{synchronous} = 1 - P_{asynchronous}$), gives information on the simultaneity perception of audio-tactile stimuli. Here the threshold values are 94.5 ms and 159 ms for AF and HF conditions, respectively. These threshold values represent the delay that will make the audio-tactile stimuli felt as synchronous, with an 86.6% probability (i.e., with a 14.4% probability of the two stimuli felt as asynchronous).

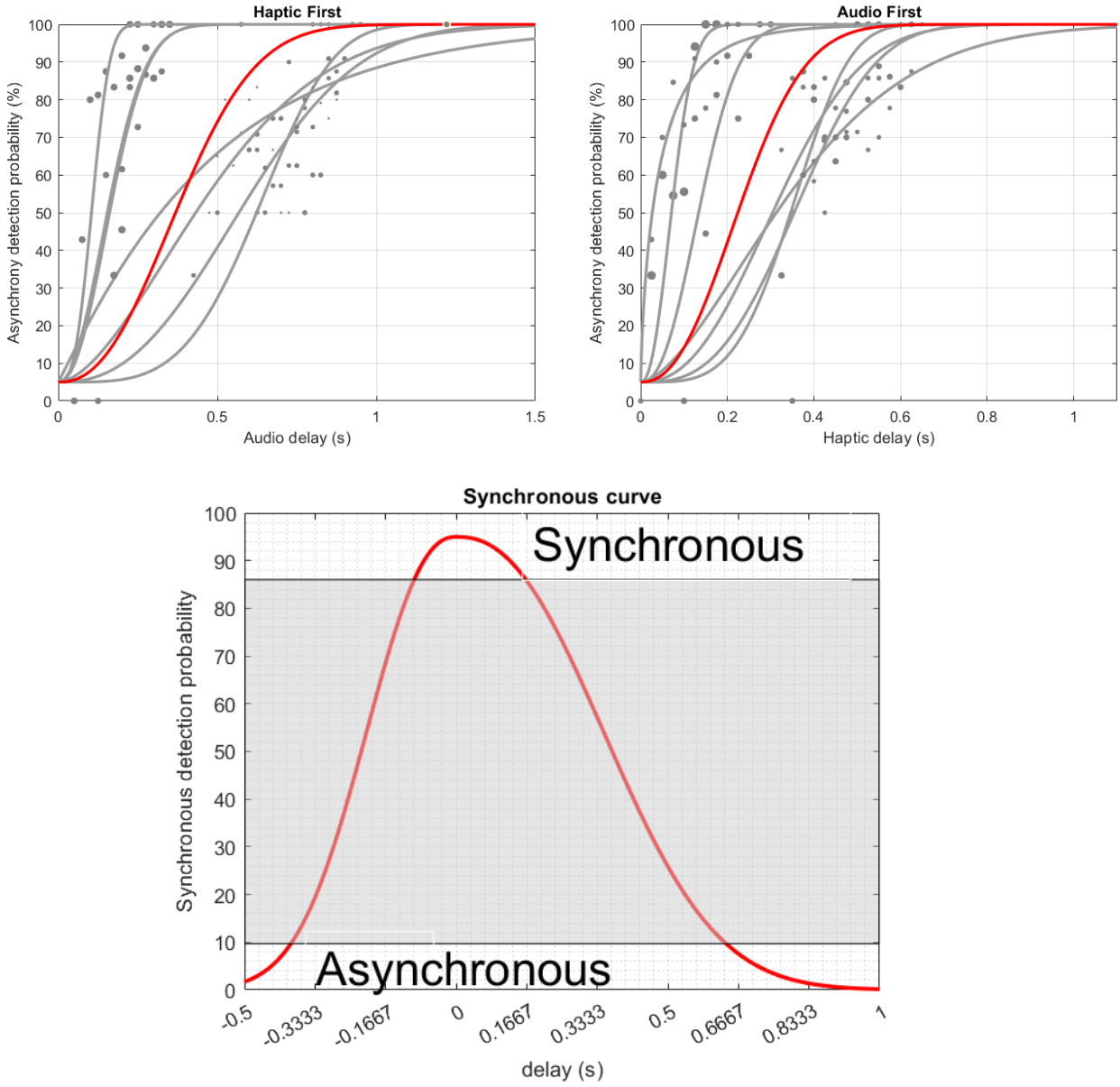

Figure 2.14: UP) Asynchronous discrimination performance in the haptic first and audio first condition for blind participants. Individual estimated curves are plotted in light gray and data (gray dots) are plotted as the proportion of trials in which the two stimuli were judged as asynchronous as a function of the audio delay. The dimension of the dots is related to the occurrence of the associated delay. The psychometric curve constructed with the most likely parameter's value (medians) is represented in red and a higher value corresponds to a higher probability of asynchronous. **Down**) Synchronous curve obtained by calculating the probability of a synchronous response. Negative delay values indicate AF condition and positive values indicate HF condition. High/low values indicate a high/low probability of feeling the stimuli as synchronous.

2.4.3 Comparison between blind and sighted people

For all our participants, threshold values in the HF condition were higher than those in the AF condition as illustrated in Fig. 2.15. The tendency illustrated in Fig 2.15 is therefore similar to what we found in section 2.3.3 where the estimated threshold values in the HF condition were smaller than those in the AF condition for each participant. However, no statistically significant differences were found for the two conditions (for both groups) as we can see in Fig. 2.16. Some participants (2 for blind and 2 for sight) show close threshold values in the HF and AF conditions. However, the majority of our participants, for both sighted and blind groups, had threshold value in the AF condition that was lower by more than 100 ms if compared to the HF condition. Some participants showed threshold values that were 300 ms to 400 ms higher with respect to the HF condition, highlighting a tendency of individuals to be more stringent (i.e., more sensitive) when the delay was on the haptic modality.

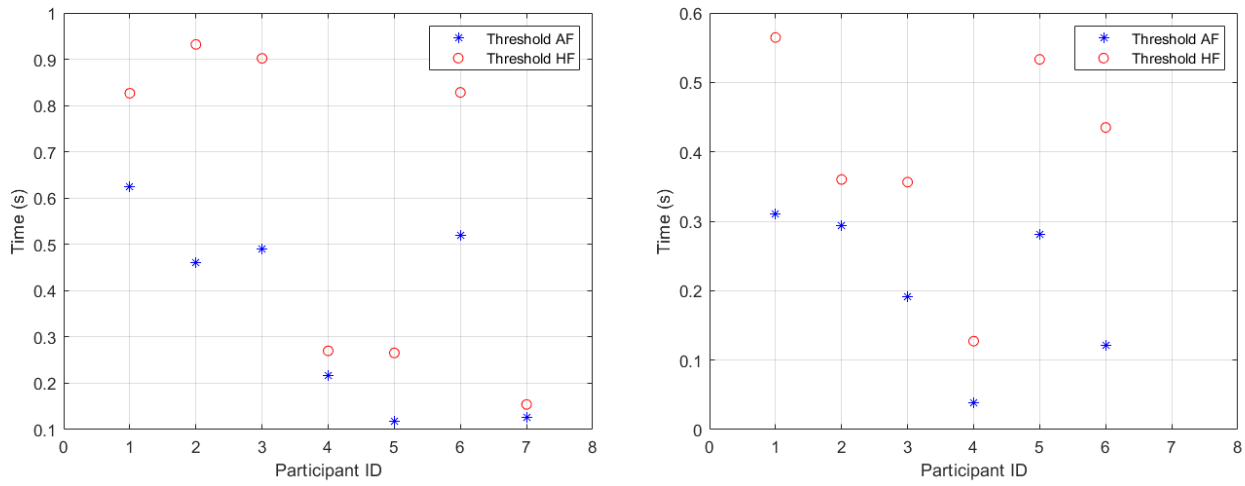

Figure 2.15: Comparison between threshold values estimated in both HF and AF conditions. **Left**) Blind participants. **Right**) Sighted participants

We believe that the number of participant in this experiment (7 for blinds and 6 for sighted) was relatively low, indicating a potential limitation of this experiment. Indeed, increasing the number of participants may improve our statistical analysis and revel a similar results as in section 2.3.3 were we have found a statistical difference between the AF and HF condition.

Overall, the estimated values of threshold and slope were higher for blind participants compared to the sighed.

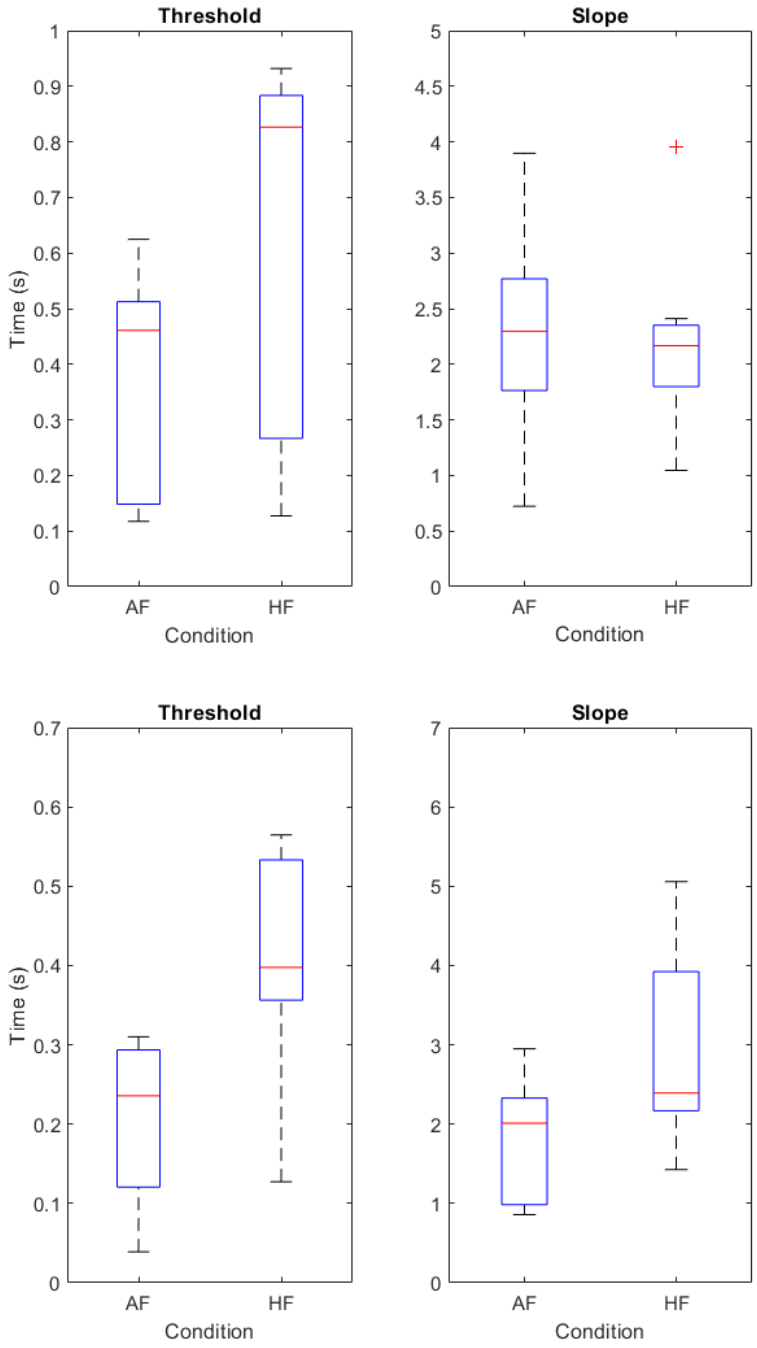

Figure 2.16: Box-plot of the estimated threshold and slope for both groups and HF and AF conditions. **Up**) Blind participants. **Down**) Sighted participants

For completeness, in Figure 2.17 we illustrate the synchronous curve for both blind and sighted

participants. From the graph, we can see a clear difference between the two populations, with sighted having a steeper slope and lower threshold values compared to their counterpart. This figure is useful as it allows us to show in a clear manner and with a glance the tendency of blind individuals to be more tolerant of delays as well as the ability of the sighted to be more sensitive to asynchronies.

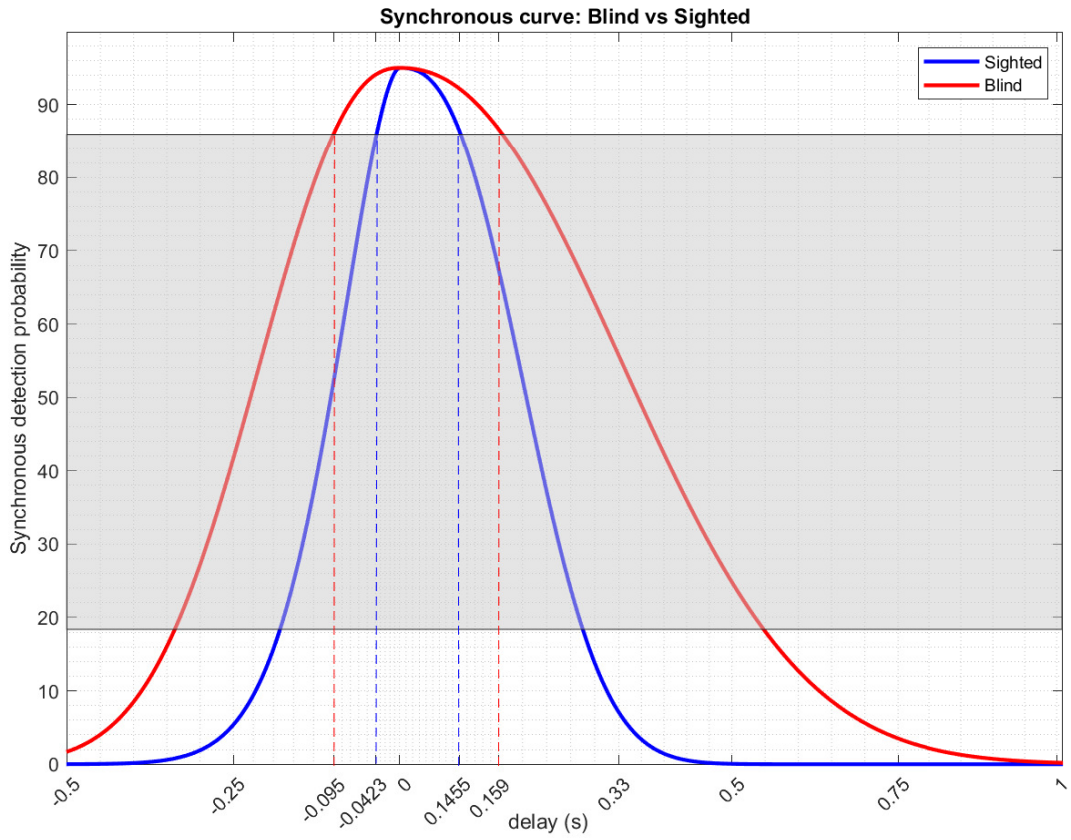

Figure 2.17: Comparison of Synchronous curve for blind (blue) and sighted (red).

2.5 Conclusion

In this chapter, we have seen three experimental investigations aiming to unveil and understand the sensory interaction between audio and tactile in active touch conditions in the specific domain of sensory synchronization.

In the first experiment, we show that sensitivity to spatiotemporal audio-tactile delays can be investigated using a novel paradigm.

Overall, for developing audio-tactile haptic technologies, a delay of about $200\ ms$ between tactile and auditory signals should be acceptable for users to experience synchronicity between the two modalities. These results suggest that a visual cue overlapping with the tactile feedback does not always influence the participant's detection of asynchronies. Surprisingly, we also saw that the finger velocity does not influence the estimated threshold, but we believe that the self-generated movement of sliding has an influence on the estimation. Indeed, it has been shown in the literature that voluntary body movement affects the sensitivity to asynchronies, leading to a phenomenon known as chronostasis, an illusory extension of perceived duration [182].

In our second study, we performed an experiment to estimate the most likely parameters for constructing the psychometric function related to audio-tactile asynchrony perception while executing a click gesture with a virtual button.

Synchronicity is a critical element of a multimodal experience. For engineers designing virtual buttons, strictly synchronous modalities are expensive to afford. Fortunately, our results demonstrate that people tolerate significant delay when haptic occurs first and a delay between the stimuli below $109.0\ ms$ ensure synchronicity. However, we also show that audio occurring first should be prohibited because people are stringent and intolerant of that condition. If this is not possible, designers should ensure a minimum delay value below $40\ ms$.

In the final study within this chapter, we conducted an experiment involving sighted participants and individuals with visual impairments. For both populations, we meticulously examined two distinct conditions (referred to as AF and HF) to determine the optimal parameters for constructing the psychometric function associated with the perception of audio-tactile asynchrony. This was specifically designed to investigate how participants perceive asynchrony when executing a click gesture with a virtual button.

Our findings reveal intriguing insights into the tolerance levels of individuals when it comes to the temporal sequencing of haptic and auditory stimuli. It appears that a noticeable delay in the audio feedback, with the haptic sensation occurring before the auditory cue, can be accepted by individuals,

provided that the delay in the auditory modality remains below a threshold of $145.5ms$. However, our study underscores a crucial point: the reverse scenario, where audio precedes haptic feedback, is met with strict intolerance among participants. In cases where this precedence is unavoidable, designers should aim to minimize the delay to a value below $42.3ms$ to maintain user satisfaction.

Furthermore, our research reveals an intriguing contrast between sighted and blind individuals. While adhering to delay limits is essential for the sighted population, it's worth noting that blind individuals exhibit a higher tolerance for delays. This observation underscores the importance of inclusivity in hardware development, especially concerning virtual buttons and interfaces.

Comparing the results of the sighted population with those in the experiment presented in section 2.3, we find consistency on individuals sensitivity relative to the stimuli condition (AF or HF). In particular, even if we used different devices, participants in both experiments were more sensitive to the condition where delays were injected in the haptic rather then in the auditory modality. Moreover, the results show a difference of 0.6% in the AF condition and a difference of about 25% in the HF condition. This indicate that while the sensitivity to stimuli order is consistent, the estimated threshold values seems to be affected by the device in use.

In summary, these findings carry significant implications for the development of hardware, particularly in the context of virtual buttons or when sliding a finger over a vibrating surface, emphasizing the importance of respecting delay thresholds to enhance user experience across diverse user groups. With our results, we aim to help the community of designers and developers to cast light on some temporal aspects of audio-tactile interaction occurring when performing a keyclick with virtual buttons.

2.6 General Guidelines

This section summarizes the main results drawn from this chapter. In particular, we aim to provide practical guidelines for developers and designers. For more details, we invite the reader to refer to Chapter 2.

Practical Guidelines

Audio-Tactile synchronization when sliding a finger (section 2.2):

- In general, synchronicity is ensured with a delay ≤ 200 ms.

Audio-Tactile synchronization with a virtual button click (section 2.3 and 2.4):

- In general, the audio first condition should be avoided; otherwise, a delay ≤ 40 ms ensures synchronicity.
- In the haptic first condition a delay value $\leq\ 109ms$ ensures synchronicity
- Caution should be paid to the stimuli duration as a shift in the point of subjective simultaneity may occur.
- Respecting these limits includes also blinds in the design.

CHAPTER 3

CROSS-MODAL INTERACTION OF STEREOSCOPY, SURFACE DEFORMATION AND TACTILE FEEDBACK ON THE PERCEPTION OF TEXTURE ROUGHNESS

3.1 Introduction

In this chapter, we begin on the exploration of the intricate realm of visual-tactile interactions. We start with an examination of the fascinating ways in which our perception can be profoundly shaped by the visual modality, with a particular emphasis on its interactions with 3D depth cues. We examine the concept of co-localized devices and their pivotal role in this perceptual interplay, exploring how touch can alter our sense of size and how the visual modality can influence our perception of attributes such as softness and roughness.

Moving forward, we cast a spotlight on various interfaces that integrate visual and tactile feedback, with a special focus on the immersive realms of stereoscopy found in virtual reality (VR) and augmented reality (AR). We uncover the synergies and possibilities that arise when these modalities converge, expanding our understanding on how our final percept is shaped.

To illustrate these concepts in practice, we present an experiment designed to analyze the impact of visual depth cues, surface deformation, and tactile feedback on the perception of texture roughness

during active touch. The findings of this study shed light on the intricate dynamics of multisensory perception.Indeed, the perception of haptic feedback is influenced by other modalities, visual and auditory, making it possible to reinforce or enrich it. However, the effect of visual depth cues, such as stereoscopic rendering and surface deformation, on the tactile perception of textures has not been studied yet, especially in an active touch condition. In this work, we investigate the perceptual interaction between stereoscopy, surface deformation, and haptic feedback in the condition of active touch implemented using friction modulation based on ultrasonic vibrations. The experimental study is based on a Visual-Tactile exploration of a virtual texture. Our objective is to understand the interaction of one modality over the other for roughness and depth perception.

This chapter has been published as a conference paper [183] and provides a comprehensive and insightful perspective on the subject.

3.1.1 Background of the proposed study

Interaction with digital content has benefited from the development of auditory, visual and haptic feedback technologies, which can make this interaction more efficient, more immersive and perceptually richer. Moreover, the combination of multiple modalities has long been an essential topic in Human-Computer Interaction, as demonstrated by early research [184] and their continuation in the more recent years [94, 185]. More specifically, multimodal feedback combining haptic and visual modalities has been shown to improve and enrich interfaces. For instance, this combination improves the performance of both 3D [186] [187] and 2D interaction techniques [188]. In the case of touch interfaces, it also offers opportunities for the visual-haptic exploration of virtual textures, which is an essential component in fields such as medical imaging , artistic expression, cultural heritage, or retail. Indeed, visual-tactile interactions may allow scientific data exploration for medical purposes, enrich artistic or historical expositions, and can be used to allow consumers to interact with products such as "touching" a virtual object before buying it.

In this context, multiple technologies have emerged that allow the rendering of haptic textures on flat surfaces, e.g. touchscreens. In particular, ultrasonic vibrations can be used to modulate the interactive forces between the finger and the surface (i.e. friction) in active touch condition. According to the neuroscientific definition, active touch refers to the act of touching and implies voluntary,

self-generated movement [30] and it is therefore suited in the case of virtual texture exploration. On the visual side, 3D displays have been developed which simulate a number of visual depth cues, both monocular (perspective, shadows, relative size, motion parallax ...) and binocular, allowing to render virtual textures with complex geometries on flat screens. In particular, they enable the separation of these cues so that their effect on texture perception can be evaluated. Combining such technologies therefore opens many opportunities for rendering and interacting with virtual content, which in turn makes it essential to understand the perceptual interaction between the visual and haptic modalities. Research on that topic has already led to many insights [116] [117] which can help design richer visual-tactile interactions.
However, there has not been research on the specific interactions between 3D visual depth cues(in AR) and tactile feedback in the context of texture roughness perception. For this, we need to understand how from the perception point of view these two modalities interact with each other. In particular, this is important because it can change the user experience in an unpredictable way. Indeed if one wants to add visual cues to an haptic display, one can end-up with skewed results where the designed experience is different from the perceived experience.

In this study we want to address the absence of research that associates haptic feedback, in active touch, with 3D visual depth cues in AR. In particular, we will focus on the interaction between AR stereoscopy and visual deformation on one side and tactile roughness on the other side. Our objective is to understand how these perceptual parameters affect the overall perceived roughness of a virtual texture rendered on a touchscreen, but also how they affect separately the perceived visual and tactile roughness.

By this investigation, we aim at improving the design of 3D user interfaces with richer feedback, especially when rendering visual-tactile textures represents a key element of the interface. We think this investigation can : 1) enrich the understanding of tactile and visual perception, and their interaction 2) inform the design of novel haptic-visual display technologies, such as mobile devices with 3D displays, or touch controllers in Mixed or Virtual Reality 3) open opportunities for new interactive applications such as 3D medical data exploration or rendering and designing virtual textures and objects.

3.2 Related Work

In this section, we review the literature on visual-tactile perception and on interfaces which combine 3D displays and haptic feedback.

3.2.1 Visual-Tactile Perception

The integration of visual and tactile modalities represents a consequence of the natural way human senses interact together. Indeed, a key connection between vision and touch is represented by the shared spatial component, present in both modalities[185]. Therefore, it has been shown that haptic information does not need to be encoded into visual information and has direct access to spatial processing [189]. This indeed represents a key element of visual-tactile interaction, highlighting the importance of physically co-located interactions. In such context, Olsson et al. [190] investigated haptic interactions with 3D displays. In their study, they highlighted the importance of having visual-haptic co-located (aligning visual and haptic) workspaces, as they do in the real world. Also Kervegant et al. [63] reported a similar result about co-location. They combined a mixed-reality headset with an ultrasound array mid-air haptic device. Their system adds tangible (tactile) feedback to virtual content and the authors highlighted how the co-location of feedback drastically enhances the presence of the object itself. Moreover, the relation between vision and haptic has been shown to be important in multimodal integration already in the early stage of human development. In particular, it has been demonstrated that touch educates and calibrates our visual perception[116] [117]. Also Picard et al. [191] reported a partial equivalence between vision and touch in a matching task related to texture perception. Even closer to our interest, Bergmann et al. [192] confirmed the equivalence of visual and tactile modalities in the perception of texture roughness.

Visual performance may also change depending on the type of tactile interaction, passive or active.

Doorn et al. [193], as an example, performed an investigation over the perception of shapes and size with vision and touch in both passive and active conditions. Their results have shown that visual influence on size judgment was greater than the influence of haptic when passive touch was involved. However, when tactile information was allowed during the active exploration, size judgment was more influenced by the haptic modality. This result shows the importance of the active touch condition in

the context of visual-tactile explorations.

Other research have demonstrated the potential influence of the visual modality over various aspects of haptic perception such as stiffness[194], compliance[195] and roughness [42]. In the more precise subject of texture perception, much research has been conducted on the separate perceptual cues for visual and tactile but also on how these modalities interact (see Klatzky and Lederman [196] for a more detailed review), especially in the context of roughness perception. Research suggested that visual and tactile perceived roughness vary similarly with respect to physical roughness [192, 197]. However, some research has shown a prevalence of tactile over visual [126] when both stimuli are simultaneously available. Perceived roughness is also influenced by many visual and tactile factors, which can interact with each other. For instance, Ho et al. [198] suggest in their study that participants judged visual roughness according to their perception of shadows, which highlight the potential tactile depth of the texture, and not using binocular cues. Visual and tactile perceptual cues can also influence one another with respect to roughness. For instance, it has been reported that colors affect the perceived tactile roughness of surfaces[199]. This influence can also be seen in the use of pseudo-haptics techniques. We can define pseudo-haptics as the phenomenon that occurs when users experience haptic feedback by observing a visual stimulus that is designed to distort depending on user input [200]. In particular, Ujitoko et al. [201] showed that it is possible to modulate fine roughness perception of vibrotactile textured surface using pseudo-haptic effect. Ujitoko [108] shows that modulating the oscillation of the visual cue (the mouse pointer in his case) makes the user feel the vibrotactile surface more uneven. Indeed, their results suggest that the larger size of the visual oscillation enlarged the perceived vibrotactile amplitude of the signal wave. In other words, the visual oscillation presented during the experiment increased the perceived roughness. Günther et al. [202] also demonstrate that cross-modality allows for limiting the variety of haptic roughness levels (on physical textures) to produce a range of perceived roughness when combined with visual feedback in virtual reality.

Previous research therefore suggests a somewhat equal role of visual and tactile modalities in the perception of roughness, the predominance of certain cues and the existence of cross-modal interactions, especially the effect of visual over tactile perception.

With the development of 3D displays where depth cues can be separately controlled, it is essential

to follow up on previous research and investigate how these specific cues influence roughness perception and how they interact with tactile perception.

3.2.2 Interfaces

Interfaces coupling visual feedback and tactile feedback (in passive touch condition) have become widespread in research in the past decade thanks to the increase in available technologies for both visual (such as stereoscopic displays and AR/VR headsets) and tactile feedback(vibrotactile, shape-changing displays, electrovibration, mid-air haptics, etc.). Many research coupling these two modalities focus on understanding the possible interactions as well as modulation of perceived properties of materials such as stiffness, softness, compliance, etc. As an example, Punpongsanon et al.[203] used a camera and a projector to visually manipulate the sense of softness perceived by a user touching (pushing) a soft physical object. The authors also added a surface deformation effect and a body appearance effect to overcome the limitations of projection-based approaches. As a result, the perceived softness was manipulated by the system such that participants perceived significantly greater softness than the actual softness. Such augmentation of the perceived softness is interesting as the system only manipulates the visual feedback, leaving unaltered the tactile side (pseudo-haptic effect).

Other research focus on the manipulation of the perceived stiffness. Yuki et al. [194], for example, developed a method to influence the perception of perceived stiffness of an object by using visual deformation of a virtual hand (instead of using object deformation). The authors show that the effect of modifying the perceived object stiffness using this method was effective. Indeed, their results suggest a stiffness increase of 1.6 times more than that with only modifying the degree of dent of virtual object.

For what regards tactile compliance perception, Kidal et al.[195] used a technique called 3D-Press to give the illusion of tactile compliance. Then, the authors added visual feedback on a screen to reinforce such tactile illusion by synchronizing the modalities effects (when compliance is tactically felt, it is also seen). Their results seems to confirm again the importance of stimuli co-location but also, they highlight the relevance of stimuli synchronization. Indeed, the temporal component of stimuli presentation is also a key element of different interfaces. For instance, Romanus et al. [62] coupled a mixed reality head-mounted display, an ultrasound array haptic device, and a smartwatch

to measure the heart rate. Their demonstration illustrates the possibility to temporally map the heartbeat (measured and displayed thanks to the AR glasses) to the haptic feedback. The update of the heart rate dynamically changes the haptic feedback and the animation, creating an enriched experience. Similarly, Han et al. [204] also coupled stereoscopic vision and haptic feedback. The main components of their interface are a stereoscopic system with beam-splitters and a thimble-formed pneumatic balloon display. Their system provides the illusion of touching 3D virtual contents while interacting directly and intuitively with these virtual objects.

To our knowledge, however, no research has been carried out on interfaces that combine co-located tactile feedback in active touch (ultrasonic friction modulation) with an AR stereoscopic display.

3.2.3 Contribution

In this work, we study the effect of 3D visual feedback in AR and tactile feedback on the perception of texture roughness in an active touch condition. We provide insights on the perceptual interaction between stereoscopy, visual surface deformation, and tactile roughness. We study their effect on the perceived visual, tactile and overall roughness of a 3D virtual texture displayed on a touchscreen. We believe that these results, by deepening our knowledge of human perception, can help the HCI community achieve a more precise and controlled user experience when designing visual-tactile interfaces. We also propose potential applications to illustrate how our results can be used in practice.

3.3 Method

The focus of this experiment is to evaluate the possible interactions between tactile feedback and 3D AR visual feedback in an active touch condition. The performed task is a visual-tactile exploration of a synthetic textured surface.

The experimental application, written using Godot Engine (v3.4), is running on a host PC with a Debian Linux OS (3.8 GB of RAM, NVidia Quadro 4000 graphics card). The application controls both the tactile and visual rendering of the textures.

3.3.1 Stimuli

Tactile

The haptic feedback used in this experimental study belongs to a haptic category known as Surface Haptics (SH). From a technological point of view, SH aims at displaying tactile feedback to the users by modulating the interaction forces between the finger and the touched surface [46]. Indeed, when a finger slides over a plate vibrating at ultrasonic frequencies, the relative friction between the finger and the surface is reduced. Such friction reduction is function of the vibration amplitude and, in particular, it decreases as the vibration amplitude increases[158]. If amplitude modulation is applied to the vibration, friction is also modulated, as are the interaction forces, as if a textured surface is touched by the user.

In this experiment, the tactile stimulator device is a glass plate resonating at 60 kHz thanks to piezo-ceramic actuators glued to it. A microcontroller manages a power unit that shapes the voltage applied to the actuators. By controlling the input signal of the plate it is possible to render virtual tactile textures. More information about the device specification can be found in [70], [156]. Moreover, a closed control-loop running in the microcontroller ensure the vibration amplitude of the plate to be as the input signal. The choice of the rendered tactile textures is based on a tactile semantic investigation performed by Dariosecq et al [205]. Their results show that the amplitude and the waveform type of the amplitude modulation signal play an important role in perceiving a texture as smooth or rough. Moreover, they illustrate that spatial period is a possible modulator for different degrees of roughness or smoothness. Starting from the adjective of the textures illustrated, we used the parameters depicted in Table 3.1.

For our experiment, we decided to use two textures, belonging to the clusters of textures associated respectively the adjectives *rough* (gray) and *smooth* (cyan), using the tactile parameters provided in Table 3.1, which correspond to one texture in each cluster.

For this reason, we decided to use a spatial period of 5000 μm for the selected textures while varying waveform (sine wave for smooth, and square wave for rough) amplitude (40 for rough, and 10 for smooth), and ratio (50%, only for square wave). One must note that because we are using synthetic tactile textures rendered through friction modulation, it is not possible to control the actual height of the texture and to provide an absolute measure of roughness, contrary to related research

employing physical textures.

Visual feedback: stereoscopy

To introduce the illusion of visual depth in the rendering of textures, we used a 27" active (quad-buffer) stereoscopic display with a pair of NVIDIA 3D Vision glasses. The glasses are shuttered at 120 Hz frequency, updating each eye 60 times per second. The images are rendered in the portion of the screen (178x102cm) below the tactile plate, as shown in Figure 3.1. A plugin was developed for the Godot engine to enable quad buffer stereoscopy to be rendered with the NVidia Quadro GPU.

Table 3.1: Tactile perception adjectives and possible corresponding synthetic texture parameters [205]. Gray and cyan rows represents adjectives related to the *rough* and *smooth* cluster respectively.

Adjective	Waveform	Period (μm)	Amplitude
granulous	rect	5000	40
rough	rect	2500	40
sandy	rect	1000	40
smooth	sin	10000	10
delicate	sin	1000	10

The visual texture was chosen to be a synthetic abstract texture, to avoid bias due to realistic or known textures which might influence the roughness perception. In all conditions, the texture was a discretized gradient noise with the same spatial frequency, chosen during preliminary tests so that the variations in luminance/depth are large enough to be perceivable but not too large, i.e. less than the fingertip size ($<$1cm), so that the participants did not expect bumps or holes in their tactile exploration. We, however, varied the seed randomly between conditions to generate variability, i.e. participants saw a different texture each time while preserving the overall visual roughness. Therefore we estimate that the visual roughness did not change among the trials due to the texture itself, but only because of the stereoscopy.

To elicit depth perception, the texture was rendered on a plane mesh and a vertex shader displaced the vertices along the Y axis (below the screen surface) according to the texture pixel value (1/white being on the surface and 0/black at 5mm below the surface).

Preliminary tests were performed to choose a texture depth (distance between dots displayed on the screen surface and dots below the surface) which was large enough for the stereoscopic effect to be perceivable but small enough to preserve the overall appearance of the texture between stereoscopic and monoscopic conditions. Finally, a depth of 5mm was found to be a good compromise. These values could be considered too large when compared with a typical profilometer measurement, such as sandpaper, where the diameter of the particles is around 500 μm [206]. However, in our work, we are not interested in displaying real textures with their proper congruent physical depth. Indeed, the choice of a synthetic abstract texture does not imply any constraints related to the texture depth, allowing designers to use larger depths than real textures profilometer measurements. In the experiment, we varied between the presence and absence of stereoscopy. When deactivating stereoscopy, the same image was displayed for both eyes, at a center position between the two eyes. Because we were interested in the effect of stereoscopy but not in an accurate perception of the depth of the texture, we chose a fixed interocular distance of 6cm for all participants and a fixed head position at 20cm above the screen. The two conditions are illustrated in Figure 3.2, with E and F showing how the texture is displaced below the surface of the screen when stereoscopy is activated, while in C and D all texture dots are displayed at the same level on the screen surface (i.e. monoscopic condition).

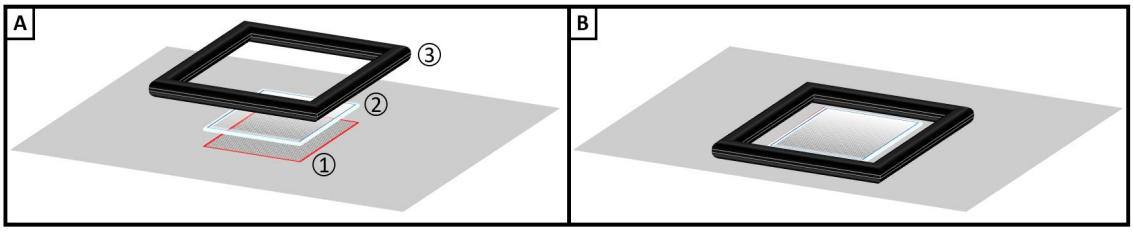

Figure 3.1: Experimental Device. **A)** Layers composing the device. **A1**-Stereoscopic screen used during the experiment where the perimeter in red highlights the used portion of the screen. **A2**-Glass plate for tactile feedback delivery. **A3**-Infrared framed used for finger tracking. **B)** How the device appears once composed. Here Power source, control, and communication are not illustrated. More information can be found in [156]

Visual feedback: surface deformation

In addition to stereoscopic depth, a visual deformation was adopted in the experimental design. Different research in pseudo-haptics has adopted surface deformation in order to reinforce or augment stiffness or elasticity properties of the rendered (real) textures. Argelaguet et al [207], such an example, used a pseudo-haptic feedback technique in order to enable the perception of local elasticity of real textures images (without any haptic device). Similarly, Kawabe [208] used a pseudo-haptic technique to modulate the perception of stiffness of objects displayed in a screen while the user, set in front of the screen, was performing gestures in mid-air without any tactile feedback. However, among different researches performed using surface deformation, to our knowledge, no research has been conducted on the possible influence produced by different deformation's shapes. Therefore, considering the type of deformation used in different researches, such as [209][210][207], and considering the neutral texture we chose for our experiment, a concave surface deformation was selected for our scope. The deformation has a sinusoidal shape (half-wave) and has been included to investigate the influence of tactile over the visual perception of deformation's depth as well as the influence of the deformation over tactile roughness perception. The deformation appears under the participant's fingertip. The deformation is synchronized with the position of the participant's fingertip and moves accordingly with the finger trajectory performing the texture exploration. The tracking of the finger is performed by an IR (Infrared) frame that surrounds the tactile plate and which exhibits a sampling frequency of 125 Hz. The finger position is indeed sent every 8ms to the OS. In the experiment, the deformation varied between no deformation (the surface remains flat below the finger) and a deformation of 10mm below the finger. Such value was chosen after pre-trial tests. We aimed to find the minimum value which could be perceived in both monoscopic and stereoscopic conditions, i.e. sufficiently deep to generate a visually perceivable change in the texture.

The shape of the deformation (see Figure 3.2) was chosen to be not related to a specific elastic behavior of a real texture since we chose a synthetic abstract (neutral) texture as a visual cue. Even if the type of deformation may have itself an effect on the perceived tactile roughness, we counterbalance this possible effect by using the same deformation shape in the trials where the condition of deformation was present. Moreover, we decided to first investigate the influence of a non-complex deformation shape, having in mind to investigate possible influences of different deformation

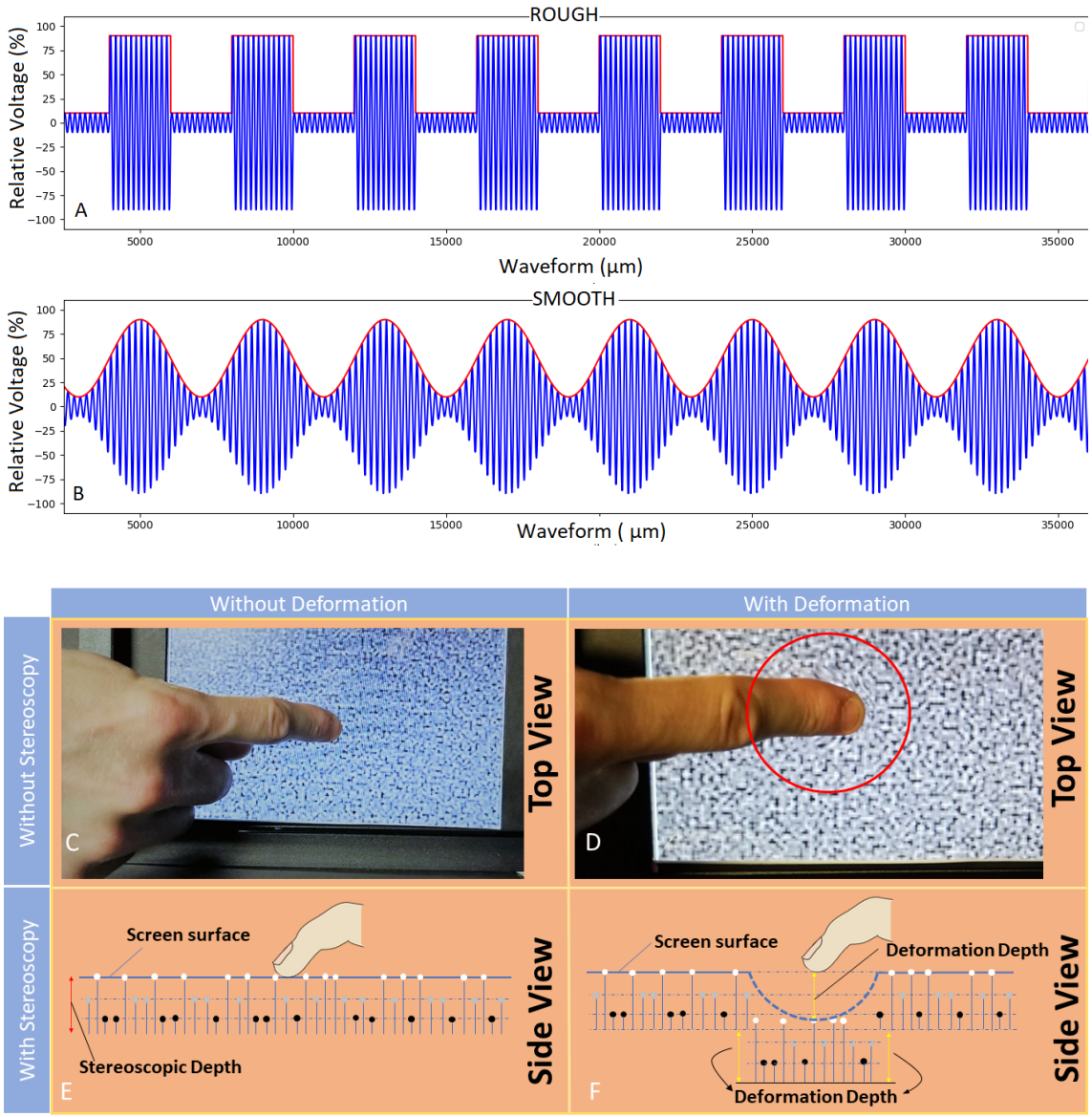

Figure 3.2: Representations of the experimental conditions: **A,B**) Tactile Texture waveform samples resulting from the parameters selected in Table 3.1. Test conditions for Texture Deformation and Stereoscopic Vision: **C**) No Deformation, No Stereoscopy. **D**)With Deformation, No Stereoscopy. **E**) No Deformation, With Stereoscopy. **F**) With Deformation, With Stereoscopy. As quad-buffer stereoscopy would not be perceivable in a picture, we chose to represent a side-view of the texture displacement when is active. Screen surface and finger in Figure E and F are not in scale

shapes in future works.
We chose the boundaries of 0mm and 10mm since our interest lies in the difference in deformation (Perceived deformation / Rendered deformation) more than in the perception of absolute values of deformation's depth.

We then chose a fixed radius of 5mm for the deformation. Such value was chosen in order to allow participants to see the deformation around their finger, avoiding occlusion of the deformation due to the finger size.
Finally, Figure 3.2 illustrates the two conditions for deformation depth. In particular, Figure 3.2 D shows the deformation when stereoscopy is not active and Figure 3.2 F illustrates how the dots moves vertically when the finger is placed on the surface and stereoscopy is active.

3.3.2 Hypotheses

In this experiment, we want to study the interaction between 3D visual depth cues (stereoscopy and surface deformation below the finger) and tactile roughness on the perception of virtual texture roughness. Based on previous research, we consider 3 hypotheses.

- H1: 3D visual depth cues in the form of stereoscopy or surface deformation will influence our perception of tactile roughness
- H2: Tactile roughness will influence our perception of visual roughness
- H3: Both 3D visual depth cues and tactile roughness will influence our perception of the overall roughness

We separate the judgement of tactile and visual roughness perception from the overall roughness perception because we believe that one modality can take precedence over the other for overall roughness. Therefore we want to isolate their effect on the perceived roughness for each modality.
To test these hypotheses we designed an experiment where we displayed a co-located visual and haptic virtual texture. We then varied the levels of tactile roughness, the depth of surface deformation below the finger, and the presence of stereoscopy, as described in sections 3.3.1.

3.3.3 Experimental Protocol

The experiment used a $2x2x2$ within-subjects design for the factors: *Texture* (type of tactile texture : Rough or Smooth), *Visual* (with Stereoscopy or without Stereoscopy), and *Deformation* (Absent: 0mm; Present: 10mm).

Participants

A sample of 22 participants (16 males, 6 females, aged M=26.68 SD=4.01) was recruited for this experimental study. The original group was composed of 24 people but two were excluded as they did not pass the stereoscopic test. 10 out of 22 were using glasses or eye contact lenses. However, 18 participants stated to have good to excellent eyesight. 21 participants were right-handed and only 1 was left-handed. Their experience with tactile devices (M=3.04, SD=1.39) and Stereoscopic displays (M=2.68, SD=1.14) was on average competent on a five-level scale.
None of the participants were suffering from motor impairment, numbness, or stereoscopic blindness. All participants took part freely in the experiment.

Exclusion criteria

Before starting the experiment, we checked that participants were outside our exclusion criteria (two in our case). First, participants were asked if they suffer from any kind of somatosensory problem, such as numbness (loss of feeling or sensation in an area of the body). This was selected since we want participants without tactile issues or tactile perception limitations. Second, we tested the stereopsis of participants using a TNO test reimplemented for our quad-buffer stereoscopic display. One butterfly (see Figure 3.3, top-left corner of the tactile display) was rendered over a circular dark background while a second one (bottom-right of the tactile display) was rendered against a random dots background. The first butterfly was easy to see (even in monoscopic condition) while the second one was more difficult to perceive, requiring a higher stereoscopic ability. Only people who were able to see both butterflies were accepted for the experiment. This second exclusion criterion was important to be tested since around 32% of the population have moderate to poor stereoscopic ability [211].

3.3.4 Experimental setup

A schematic view of the device we used for the experiment can be seen in Figure 3.1. It employs a stereoscopic screen (A1, with associated NVIDIA 3D Vision shutter glasses) on top of which is placed a glass plate actuated by piezoceramic transducers, which generate the tactile feedback. Around the plate, there is an IR finger tracking frame. Parts of the stereoscopic screen are not used and are covered using a black foam board to avoid undesired visual cues. As a result the portion of the screen in use (in red in Figure 3.1 A1) is located at the center.

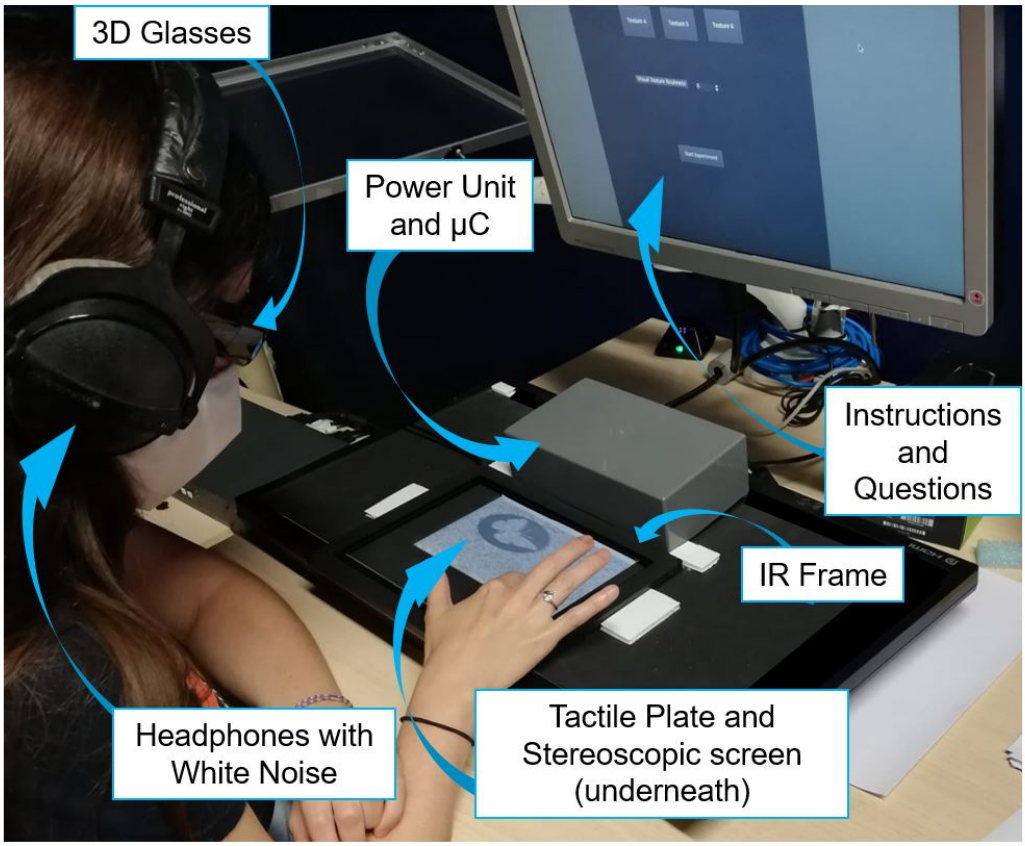

Figure 3.3: A participant during the stereoscopic TNO test (butterfly on the screen). We illustrate the experimental set-up with the different components of system detailed in the picture.

Visually textures can be displayed without or with stereoscopy, i.e., respectively with all dots on the surface (0 mm) or with some of them 5mm below. The haptic feedback is delivered through the glass plate (Figure 3.1 A2)). Thanks to the finger tracking system (Figure 3.1 A3) virtual textures can also be deformed locally, appearing as being pushed below the finger. Finally, the full setup

provides a co-localized 3D AR visual and haptic feedback.

Procedure

Participants sat in front of the setup wearing a pair of 3D glasses and headphones playing white noise to hide the audible noise generated by the vibrating tactile plate (see Figure 3.3).
After the application of the exclusion criteria, a training session was started. Participants were first presented with two tactile textures (on a blank screen) and were told that the first was at a level of 20/100 on our roughness scale, while the second was at a level of 80/100. These two textures corresponded to the two textures which were used for the smooth and rough tactile conditions, as described in Section 3.3.1. Such textures were used as a reference and participants were asked to remember them in order to easily rate texture roughness in the linear scale (0 to 100). The reference values (20 and 80) were chosen because they correspond to textures from the smooth and rough clusters in [205] and because these values leave room for lower and higher roughness ratings. Participants were then asked to visually judge a texture in roughness with a value between 0 (completely smooth) and 100 (completely rough). This value was then used as a personal baseline for visual roughness judgment during the experiment. Once again, we are interested in differences between conditions and not in absolute values of visual roughness. Therefore, the visual textures remained very similar, as explained in Section 3.3.1, so that participant's perceived roughness would change only because of the experimental conditions.

In the experimental phase, each participant therefore performed 16 visual-tactile explorations (2 *TactileFeedback* * 2 *Stereoscopy* * 2 *VisualDeformation* * 2 repetitions) for a total of 352 trials for all the participants. These trials represent all the combinations of tactile feedback (smooth or rough as described in Table 2.3), stereoscopy (presence or not) and surface deformation (presence or not). To avoid biases due to the presented order of the trials, the order of the conditions was counterbalanced between participants using a balanced Latin square algorithm.
At each trial, the visual texture as depicted in Figure 3.2 C and D appeared in front of participants. At this point, participants performed the visual-tactile exploration. They were free to explore each texture (visually and tactically) for as long as they needed, using the index of their dominant hand. The exploration was restricted to a single finger as the hardware used for this experiment does not

allow multitouch tactile feedback.

Data Analysis

We considered two repetitions for each condition were enough for participants to express their perception. Indeed, unlike pointing tasks, participants had time to explore the texture and judge their perception (approximately 2 minutes for each trial). In general, the 2 repetition for a given combination was consistent among participant.

While exploring the visual-tactile texture, 5 questions were asked to the participants in order to verify our hypotheses : the level of perceived tactile roughness (between 0 and 100), the level of visual roughness (between 0 and 100), the depth of deformation below the finger in millimeters (between 0 and 10mm), the overall perceived roughness (on a scale of 1 to 5 between very smooth and very rough) and the presence of stereoscopy (yes or no, this data was kept for further studies). For both levels of tactile and visual roughness participants rely on levels given during the training phase.

In the case of deformation depth, we asked participants to estimate depth in the continuous scale 0-10mm, although conditions were either 0mm or 10mm. We are not interested in absolute or precise values for the perceived deformation depth, but rather to detect changes due to variations of our three independent variables.

Finally, to evaluate the weight of all factors in a global roughness judgment scale, we use the overall roughness, for which we asked participant to evaluate their perception of roughness taking into account both modalities.

The experiment duration averaged around 30 minutes per participant.

3.4 Results

In this section, we present the obtained results for the dependent variables: Perceived tactile roughness (*TactileRoughness*), Perceived visual roughness (*VisualRoughness*), Perceived depth (*PerceivedDepth*), and Perceived overall roughness (*OverallRoughness*). For each, we performed a repeated-measures ANOVA for the factors *Stereoscopy* (Yes/No), *VisualDeformation* (Yes/No), *TactileFeedback* (Smooth/Rough).

Normality was tested with a Shapiro-Wilk test. For variables that did not follow a normal distribution, namely *TactileRoughness*, *PerceivedDepth* and *OverallRoughness*, we applied an Aligned Rank Transform [212] before the ANOVA. We ran the analyses with R v4.0.3.

3.4.1 Perceived tactile roughness

The perceived tactile roughness for a trial was computed as the difference between the score (0-100) given by the participant during the trial and the score associated with the presented texture (smooth=20, rough=80, given to participants during the training phase).

An ANOVA showed statistically significant main effects of *Stereoscopy* ($F(1,147) = 10.45, p = .001, \eta^2 = 0.189$), *VisualDeformation* ($F(1,147) = 8.28, p = .004, \eta^2 = 0.152$) and *TactileFeedback*($F(1,147) = 9.6, p = .002, \eta^2 = 0.173$). It also showed significant interactions *VisualDeformation*TactileFeedback*($F(1,147) = 15.09, p = .0001$) and *Stereoscopy*VisualDeformation *TactileFeedback* ($F(1,147) = 9.23, p = .002$).

Post-hoc Holm adjusted pairwise t-tests showed a number of statistically significant differences. Some seem to be supported by changes in *TactileFeedback*, such as *StereoNo- TactileRough-DeformYes* and *StereoNo-TactileSmooth-DeformYes* ($t(147) = -4.207, p = .0011$), *StereoNo-TactileRough- DeformYes*and *StereoYes-TactileSmooth-DeformYes* ($t(147) = -4.385, p = .0006$). This suggests that switching from a smooth to a rough tactile feedback might increase the perceived tactile roughness, confirming our choice of tactile parameters from the literature.

Other statistically significant differences seemed to be supported by changes in visual feedback (*Stereoscopy* and *VisualDeformation*) in the *TactileSmooth* condition, such as *StereoNo-TactileSmooth-DeformNo* and *StereoNo-TactileSmooth-DeformYes* ($t(147) = -5.275, p < .0001$), *StereoNo- TactileSmooth-DeformNo* and *StereoYes-TactileSmooth-DeformNo* ($t(147) = -4.07, p = .0018$), *StereoNo-TactileSmooth-DeformNo* and *StereoYes-TactileSmooth- DeformYes* ($t(147) = -5.45, p < .0001$). This suggests that in the case of a smooth tactile feedback, adding either stereoscopy or surface deformation below the finger increases the perceived tactile roughness, therefore confirming H1. Figure 3.4.a shows the differences in perceived tactile roughness when stereoscopy and surface deformation are enabled for smooth textures.

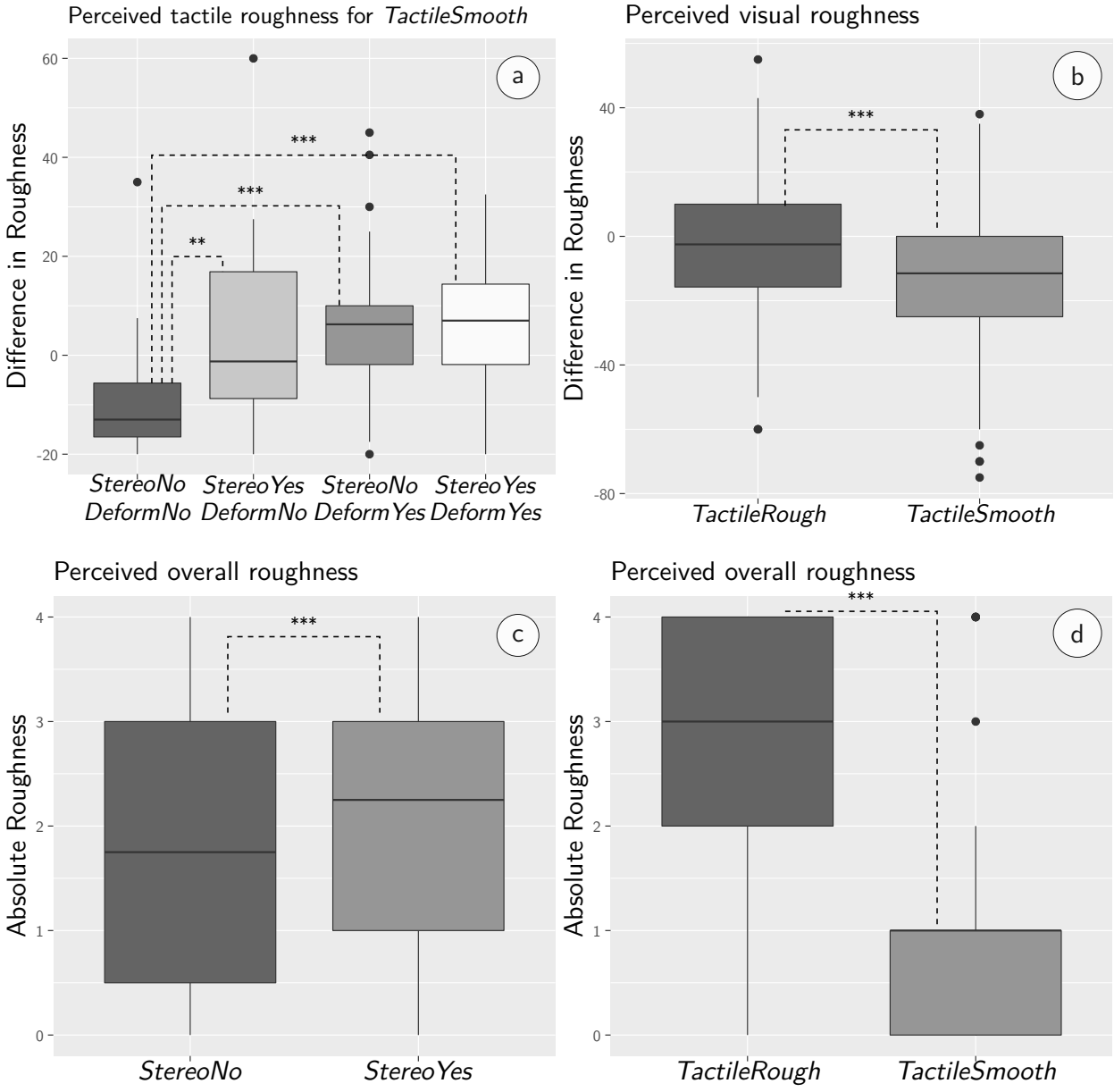

Figure 3.4: Plots of statistically significant results: a) Effect of *Stereoscopy* and *VisualDeformation* on the perceived tactile roughness, b) Effect of *TactileFeedback* on the perceived visual roughness, c) Effect of *Stereoscopy* on the overall perceived roughness, d) Effect of *TactileFeedback* on the overall perceived roughness.

3.4.2 Perceived visual roughness

The baseline roughness, i.e., without tactile feedback, was quite high ($mean = 60.68, sd = 20.74$ on a scale from 0 to 100). The perceived visual roughness for a trial was computed as the difference between the score (0-100) given by the participant during the trial and the baseline score given on the visual texture without tactile feedback or deformation during the training phase.

An ANOVA showed an interaction *Stereoscopy*VisualDeformation* ($F(1, 21) = 8.843, p = 0.007$, $\eta^2 = 0.3$) and main effects of *Stereoscopy* ($F(1, 21) = 5.12, p = 0.034, \eta^2 = 0.2$) and *TactileFeedback* ($F(1, 21) = 25.69, p < .001, \eta^2 = 0.55$). Post-hoc tests reveal that the difference of perceived visual roughness between *TactileRough* and *TactileSmooth* is statistically significant ($t(174) = 3.79, p = .0002$), with almost no reduction of perceived visual roughness from the baseline in the case of a rough tactile feedback ($mean = -2.04, sd = 17.2$) to a larger reduction in the case of smooth tactile feedback ($mean = -12.8, sd = 20.3$). This suggests that a smoother tactile feedback leads to a lower perceived visual roughness (relative to the participants judgment of the texture without any feedback), as shown on Figure 3.4.b, and therefore confirm our hypothesis H2.

3.4.3 Perceived depth

An ANOVA performed on *PerceivedDepth* revealed statistically significant main effects of *Stereoscopy* ($F(1, 147) = 13.42, p = .0003, \eta^2 = 0.077$), *VisualDeformation* ($F(1, 147) = 513.23, p < .0001, \eta^2 = 0.828$) and a significant interaction *Stereoscopy*VisualDeformation* ($F(1, 147) = 12.87,\ p = .0004$).

Post-hoc Holm adjusted t-tests showed statistically significant differences between *StereoNo- DeformNo* and *StereoNo-DeformYes* ($t(147) = -13.98, p < .0001$), between *StereoNo-DeformNo* and *StereoYes-DeformYes* ($t(147) = -15.89, p < .0001$), between *StereoNo-DeformYes* and *StereoYes-DeformNo* ($t(147) = 13.93, p < .0001$), between *StereoYes-DeformNo* and *StereoYes- DeformYes* ($t(147) = -15.84, p < .0001$).

Differences in perceived depth went from almost none when no deformation was present ($mean = 0.3mm, sd = 1.01mm$) to almost half the actual deformation size when there was deformation ($mean = 4.64, sd = 2.31$). This suggests that the difference in perceived depth are mainly due to the surface deformation, which were correctly detected by participants. However, we could not conclude on this as we did not see any statistically significant main effect of *TactileFeedback* over

the *PerceivedDepth*.

3.4.4 Perceived overall roughness

An ANOVA performed on *OverallRoughness* revealed statistically significant main effects of *Stereoscopy* ($F(1,147) = 10.53, p = .001, \eta^2 = 0.071$) and *TactileFeedback* ($F(1,147) = 283.12$, $p < .0001, \eta^2 = 0.850$). No significant interactions were observed.

This result partially confirms H3, with stereoscopy and tactile roughness both increasing the overall perceived roughness. We can however not conclude regarding the effect of surface deformation below the finger. The effect of *TactileFeedback* and *Stereoscopy* on overall perceived roughness is depicted on Figure 3.4.c and 3.4.d.

3.5 Discussion

In this section, we provide an in-depth analysis of our results. We then discuss their implications and potential applications.

3.5.1 Modification of the perceived tactile roughness

Our results suggest that tactile roughness can be amplified on smooth textures by adding 3D visual feedback, in the form of stereoscopic rendering of the texture surface and/or by the deformation of this surface below the finger. More precisely, we observe an increase of more than 15 points (on a 100 points scale) when adding 3D visual feedback. This is especially interesting in the case of devices or tactile feedback technologies where smooth tactile feedback represents a key element of the system. In these circumstances, adding 3D visual feedback can help provide an impression of rougher tactile textures.

Furthermore, some of our participants (P4, P14, P20, P21) explicitly stated that the combination of stereoscopy, surface deformation, and smooth tactile feedback was their preferred condition, with P4 describing it as a congruent experience.

While texture deformation can easily be implemented on any display through mesh deformation at the touch coordinates, stereoscopy requires additional equipment, such as glasses for active or passive

stereoscopy, or even auto-stereoscopic or multiscopic displays. However, such technologies are now widely accessible, from large screens to mobile ones.

Although the effect on feedback amplitude would need to be confirmed by further experiments, we believe this could also help increase the perceived roughness in case of weak tactile feedback.

However, we can not conclude on the preponderance of either stereoscopy or surface deformation on the perception of tactile roughness. We can neither conclude on the effect of stereoscopy nor surface deformation in the case of rough textures. We believe that this could be due to the preponderance of tactile over visual, with a strong tactile perception "taking over" the visual perception and masking its effect. Moreover, this result seems to confirm the founding of Ujitoko et al. [201] in a scenario where the presence of deformation, as pseudo-haptic feedback similar to cursor oscillations, increase the perceived roughness of textures.

In contrast with their results, in our experiment the increase of perceived roughness appears to happen only for textures defined as smooth [205]. We believe this could be due to the predominance of the tactile modality. In the case of rough textures, the added visual depth cues might not make a strong enough perceptual change, although this would need to be confirmed by further study.

3.5.2 Modification of the perceived visual roughness

Our results suggest that the perception of visual roughness can be modified by changing the roughness of the tactile feedback, with an increase of more than 10 points in our subjective 100 points scale when using a rough tactile texture. In particular, it seems that using a square waveform instead of a sine waveform, which was proven to increase the perceived roughness on textures rendered with ultrasonic friction reduction [205], increases the level of perceived visual roughness.

This can be useful in cases where one wants to modulate the perceived roughness of a displayed texture, for commercial (e.g. allowing users to experience a wider range of material on displayed products), cultural (e.g. allowing visitors to get a more accurate feeling of the roughness of the surface of exhibited artifacts) or artistic applications (e.g. providing additional perceptual cues in an art gallery by designing multimodal experience, similarly to what was done by Vi et al. [213]).

In addition, when the display resolution is too small (e.g. on mobile devices) to offer enough detail to render, tactile feedback can be used to modulate the perceived visual roughness when no more

visual details can be added.

These results are important also because they confirm the correctness of the choice we made for the synthetic texture used during the experiment. Indeed, it seems that the participant's perception changed because of either the added stereoscopic rendering or the change in tactile feedback, meaning that the variations in roughness perception were due to the conditions but not to the small variations in the randomly generated visual texture, or to the nature of the texture itself. Our results finally seem in accordance with previous research on visual cues influencing roughness perception, such as light direction which emphasizes reliefs. In fact, they suggest that stereoscopy and the induced binocular disparity could be another factor of perceived roughness on visual textures.

3.5.3 Modification of the overall roughness and preponderance of tactile feedback

In the overall estimation of roughness, although we see an effect of stereoscopy, we observe that it is mainly the tactile feedback that has an effect. Effects size indeed show that the statistically significant effect of stereoscopy remains small, while the effect of tactile roughness is very high and leads to an increase of around 2 points in our 5 points scale, representing an increase of the overall perceived roughness of about 40% in our subjective scale. Therefore, it seems that when participant have to judge the overall roughness (after the integration of visual and tactile information), the tactile information weighs more than the visual information. Indeed, this suggests that if the goal is to modulate the perception of the overall roughness on a visual-tactile texture, more relevance should be placed on tactile rendering, although stereoscopy can also increase the roughness but to a lesser extent.

This result may seem to contradict theories on multisensory cues integration [214], i.e. the visual modality should dominate the overall perceived roughness because of a lower estimation variance. However in our case the visual aspect of the texture did not change much between conditions (we used variations of the same visual noise), with only the added stereoscopic depth cue, which might explain the relative importance of haptic. We believe that by changing the visual frequency or smoothness of the texture the visual modality would dominate.

3.5.4 Limitations

Our results should be taken with precautions, given some limitations in our experiment.

First of all, we used generated noisy textures to avoid the effect of known materials on the perceived roughness. Further experiments should investigate if our results remain with known / real 3D textures, i.e. fabric, wood, sand, and others. Moreover, we used a fixed head position on our 3D rendering to isolate the effect of stereoscopy in visual depth perception. However, head-tracking would provide additional visual cues of the depth of textures and might increase the perceived roughness.

Our implementation of friction reduction by ultrasonic vibration did not take finger speed into account. Depending on the participant's finger speed during texture exploration, we may have introduced variability in the spatial period of perceived tactile textures. Even if we were interested in the effect of the tactile waveform (and not spatial period), we do not believe such (limited) variability was an issue. Indeed, it has been shown that roughness perception of macro-textures are not affected by finger velocity [215] [35] [166]. Finally, a fixed finger velocity of 60 mm/s has been set for our haptic surface and these value has been selected in a pretrial session.

We did not use absolute scales for testing visual or tactile roughness. We used differences between scores given or obtained in training and scores of the experiment. Nevertheless, relying on an absolute scale, such as physical textures, may allow more precise measurement of perceived roughness. However, the objective of the investigation was to exclude the use of real texture to avoid bias due to known textures.

The ratio of males and females in the experiment should be taken into account while considering the generality of the results. Even if we aimed to have a balanced population in terms of gender, due to the pandemic, we had to deal with a reduced number of participants. This results in a less balanced population where males represent approximately 72% of our population. Indeed, this unbalanced gender ratio may affect the generality of the insights but still, no research has been conducted on the effect of an unbalanced group for this kind of studies [216]. Indeed, no guidelines exists suggesting which ratio between male and female is appropriate or which effect an unbalanced group may have on the results.

3.5.5 Example applications

In this section, we envision three example applications of our results with friction modulation feedback and stereoscopy or surface deformation.

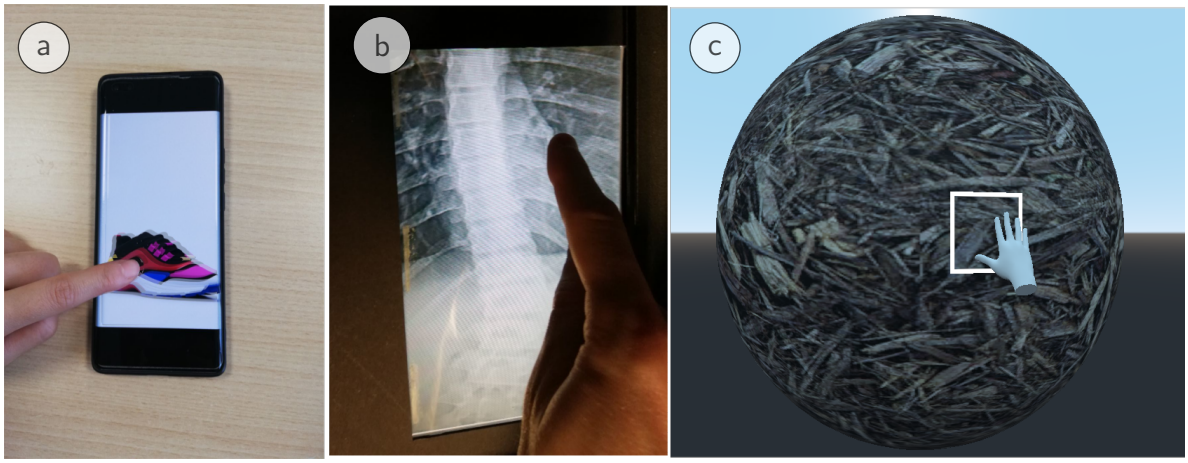

Figure 3.5: Mockup applications of combined 3D visual and tactile texture exploration : **a**) an auto-stereoscopic mobile display for visual-tactile material rendering **b**) Visual-Tactile exploration of a 3D texture in a medical context **c**) Texture rendering / editing in Virtual Reality

3.5.5.1 Tactile textures amplification on mobile devices

A first possible application would be the use of mobile devices. In this scenario, it is possible to combine tactile feedback (based on friction modulation) and stereoscopic rendering (using auto-stereoscopy) for texture exploration. This combination would allow displaying textures with a wide range of perceived roughness and would be useful in different situations, ranging from commercial products presentation to the exploration of the surfaces of exhibited artifacts in cultural heritage contexts. A mockup for the exploration of the material of clothes is shown in Figure 3.5.a.

3.5.5.2 Exploration of 3D textures

A second application that we envision is to enrich the visual examination of 3D visual textures, such as MRI scans, by adding tactile feedback to amplify the perceived visual roughness. This combination could be useful in scientific data explorations, such as large point clouds, or medical 3D textures. Figure 3.5.b depicts a chest scan displayed on our prototype, which could be explored both visually

and tactically with an increased range of roughness, therefore potentially helping the discrimination of zones with different densities in the texture.

3.5.5.3 3D objects interaction with visual-tactile surfaces

The third envisioned application involves the combination of stereoscopy, surface deformation, and tactile feedback in a virtual reality environment for the edition or exploration of 3D surfaces. As depicted in Figure 3.5.c, a virtual frame representing a physical handheld tactile plate could enable the selection and exploration of part of the 3D scene, similar to what was proposed by Montano et al. [217] for dense environments selection.

3.6 Conclusions

In this work, we investigated the interaction between 3D visual cues (stereoscopy and surface deformation) and tactile feedback and their effect on roughness perception in the context of virtual texture exploration. Adding to general knowledge on perceptual interaction and visual-tactile display, our results suggest that stereoscopy and deformation modify tactile perception in the case of smooth tactile textures, that visual perception is affected by tactile feedback and that tactile feedback is prominent in the overall judgment of roughness. Our results can be used as guidelines for all interface designers that want to use stereoscopy/deformation and tactile feedback in an active touch condition (such as ultrasonic vibration, electrovibration, or others). Indeed, this can help designers in different fields such as education, scientific data exploration, medical applications, gaming, and many others.

In future work, we want to implement stereoscopic vision using a VR headset with the addition of a head-tracking and hand-tracking system for visual-tactile interactions. In this situation, we would like to investigate the effect of a virtual hand instead of a real one in visual and tactile perception. Finally, we aim at combining auditory feedback with our existing system. We are interested in how this modality can affect our visual and/or tactile perception since this still needs to be addressed.

3.7 General Guidelines

This section summarize the main finding drawn from chapter 3 and it is aiming at providing practical guidelines for designers and developers.

Practical Guidelines

- Adding stereoscopy on a smooth texture increase the perceived tactile roughness of about 11%.
- Adding texture deformation on a smooth texture increase the perceived tactile roughness of about 20%.
- Adding stereoscopy and texture deformation on smooth texture increase the perceived tactile roughness of about 22%.

CHAPTER 4

SPATIAL HAPTIC: TACTILE FEEDBACK IN VR

4.1 Introduction

In this chapter, we explore the intricate realm of haptic feedback for guidance. Our exploration begins with an in-depth look at the pivotal role of haptic feedback within multimodal systems. The haptic channel represents an intriguing avenue for conveying information to users, although with certain challenges, such as limitations in the amount and complexity of information it can transmit.

To bring these concepts to life, we will study the realm of haptics and virtual reality (VR), with a specific focus on haptic-based directional information. Our examination will encompass an analysis of various haptic technologies and the optimal body locations for effective communication. Indeed, advancements in haptic feedback using tactile sensations have been remarkable, but the area surrounding the ears, with high haptic sensitivity, has been underexplored.

By submerging our participants in VR, this chapter explores the effectiveness of vibration headphones (VH), specifically the Razer© Kraken V3 HyperSense, in providing directional information for object spatial localization. We aim to assess participants' ability to locate non-visible objects using haptic, audio, or combined audio-haptic feedback. Our objective is to analyze user performance and, by extension, the spatial resolution of our system in providing precise directional information.

It's important to highlight that this chapter draws extensively from Brahimaj et al. [218], con-

tributing to a comprehensive understanding of the subject matter.

4.2 Related Work

The role of the haptic signal within a multimodal system can be likened to that of a member of a sensory team. Haptic signals collaborate with other senses, fulfilling two primary functions: reinforcing information for a shared perception or providing complementary details for distinct perceptions.

For instance, in a scenario where an automobile driver receives reinforcing multimodal information, various sensory cues synergize seamlessly. Visual cues from a map display visually convey an upcoming turn, auditory input provides vocal directions, and haptic feedback in the form of vibrations on the turning side of the seat or steering wheel signals the approaching turn. These modalities collectively reinforce navigational information, ensuring that the driver is well-informed and alert. Alternatively, a visual map may present a bird's-eye view of the route, while haptic feedback provides information about the distance to the turn. In this scenario, visual and haptic information complement each other by offering distinct details during the navigation task. In a navigation scenario where the driver has a general sense of their location and needs a subtle prompt to make the right turn, the low-bandwidth, low-detail haptic channel appears to be the best solution. A detailed, high-quality map would be inappropriate in this context, as it would divert the driver's attention from their primary task—driving.

Moving into the realm of virtual reality (VR), many researchers have worked with the auditory channel to provide directional information for localization or navigation tasks. VR is an incredibly immersive technology that transports users into simulated environments through the use of a specialized headset. Indeed, to render specialized sound in VR, individual or generic Head Related Transfer Functions (HRTFs) are typically employed. HRTF is a phenomenon that describes how an ear receives sound from a sound source. Berger et al. [219] demonstrated that a generic HRTF is good enough to enable a good auditory source localization. Therefore, specific HRTF based on individuals may necessary to achieve good accuracy for auditory source localization. Moreover, it was also shown that providing visual information such as hand location and room/environment dimensions led to better localization performances compared to the situation where no visual information was provided [220].

In recent years, the development of haptic feedback has further enriched VR interactions. Haptic feedback refers to the use of tactile sensations to communicate information or simulate touch in virtual environments. This aspect of VR technology has experienced significant progress, primarily due to the introduction of novel haptic devices such as haptic surfaces, haptic gloves, and wearable haptic devices [44].

The integration of haptic feedback holds great importance as it has the potential to significantly improve VR experiences, enhancing user interaction and immersion. Most importantly, designers now have the ability to not only enhance the sense of touch in virtual environments but also convey information through this novel channel, such as directional information, opening up exciting possibilities for innovative design approaches.

Researchers have proposed different multimodal systems to guide, for example, visually impaired individuals in reaching their destinations. As an example, a common approach involves using haptic directional information or audio directional cues to indicate the direction to move. A simple approach could be the one proposed by Van Erp et al. [221] that employed a belt-like device to encode distance and orientation, given a reference point to reach. Their results demonstrated that eight actuators were sufficient to provide good localization performance, with a spatial resolution of 45^o (later improved by Heuten et al. [222] to 30^o). However, the body location represents an important aspect when conveying information through the haptic channel.

Interestingly, our skin happens to be the largest sensory organ in our body, although the level of sensitivity varies across different regions [223]. Over the years, researchers have been intrigued by the potential to convey information through the sense of touch. To explore this possibility, they have investigated various body locations and experimented with different tactile signals.

In an exemplary study, Paneels et al. [224] utilized a tactile bracelet capable of generating static and dynamic patterns to convey directional information. Their research showcased that dynamic patterns were recognized with greater accuracy in comparison to static patterns, revealing the fascinating phenomenon of phantom sensation, wherein multiple simulations are perceived as a unified stimulation. This intriguing finding sheds light on the intricacies of tactile perception and carries significant implications for designing haptic feedback systems in VR.

Taking into account various tactile devices, Meier et al. [225] conducted an extensive investigation

to assess the effectiveness of three vibrotactile devices (a belt, a wristband, and shoes) in a pedestrian navigation task. Their compelling findings indicated that, in less intricate geographical situations, vibrotactile feedback alone could suffice to facilitate accurate navigation. However, in more complex scenarios, the authors propose that the integration of additional guidance mechanisms might be imperative to ensure precise and reliable navigation.

When it comes to providing haptic feedback, researchers have primarily focused on different body parts, with the hand being the most extensively studied region. The hands and face possess a higher density of tactile afferents, making them well-suited for haptic stimulation due to their heightened tactile sensitivity [28]. Dim et al. [226] investigated vibration feedback sensitivity across nine different body parts, including the ear, neck, chest, waist, wrist, hand, finger, ankle, and foot. Their study revealed that the ear, hand, and foot exhibited the highest sensitivity among the examined regions.

Consequently, technologies such as haptic gloves and hand-based devices have been developed to stimulate the fingertips, palms, and hands, offering an enhanced haptic experience. However, with the increasing popularity of head-mounted displays (HMDs), researchers have turned their attention toward exploring the possibility of providing tactile stimuli directly to the head, where the haptic systems can be co-located with the display.

One simple approach has been to integrate vibrotactile actuators into HMDs, utilizing helmets, as demonstrated by Kaul et al. [227]. Similarly, Kerdegari et al. [228] integrated seven vibrotactile actuators into a firefighter helmet's forehead area to guide users in low-visibility environments. In a navigation task, the authors compared haptic and auditory feedback and found that the haptic modality resulted in lower route deviation, highlighting its potential for improving navigation accuracy.

Focusing on the face region, another avenue of exploration involves temperature displays, leveraging the relatively high density of thermoreceptors in the forehead [229]. These technologies have shown promise in increasing immersion and providing directional cues [230]. However, the author in [230] noted a significant difference in recognition between cold and hot stimuli, rendering thermal-directional systems less suitable for navigation purposes. Rietzler et al.[231] introduced airflow from various fans as a means of providing haptic feedback, thereby enhancing the user's sense of presence and realism by adding environmental information (such as wind) to the virtual experience.

Despite numerous studies investigating haptic systems that stimulate the face, hand, wrist, fingers,

and other body regions, relatively little attention has been given to haptic feedback in the area surrounding the ears [232], which, as previously mentioned, exhibits a high degree of haptic sensitivity [28] [226].

Gil et al. [233] explored the use of a mid-air haptic device capable of providing ultrasonic tactile feedback to stimulate the user's face. They achieved a high recognition rate when targeting the center of the forehead. However, the authors opted not to use ultrasonic stimuli on the ears due to safety concerns.

Another noteworthy study by Lee et al. [234] involved the use of a compact ear-worn haptic device that could deliver information by stimulating three different locations in each ear's auricle. The authors successfully provided haptic feedback through the ears' auricles, emphasizing the potential of this approach. Nevertheless, they did not evaluate the accuracy of navigation scenarios in their research.

One promising method to convey directional information in VR is by providing haptic feedback utilizing the ear as a location on the body through the use of vibration headphones (VH). These commercially available headphones integrate vibrating motors inside each ear cup to deliver haptic feedback. This method may be a good solution for substituting directional information typically provided with auditory-based HRTF with a tactile-based HRTF. To the best of our knowledge, we are the first to investigate the application of these VH in a Virtual Environment (VE), examining their effectiveness in providing directional information in a localization task. Based on the findings in [219], we believe that a generic haptic-based HRTF may be sufficient to have comparable audio-based HRTF accuracy and comparable efficiency in providing directional information.

Through an extended exploration and adept implementation of haptic feedback within the realm of Virtual Reality (VR), our objective is to unveil novel avenues for conveying information. This haptic feedback holds the potential to create captivating virtual experiences, including applications in navigation and beyond. Thus, the central query driving our research is: Can we substitute audio-based HRTF with its haptic-based counterpart for object localization, all while upholding accuracy? This inquiry is rooted in a comparative analysis of their performance.

4.3 Vibration Headphones' mechanical characterization

Prior to utilizing the VH (vibration headphone), it is essential to characterize its mechanical behavior. To achieve this, we employed a sweep frequency signal spanning the range of 1 Hz to 200 Hz as the input stimulus. The vibrations generated by the VH in response to this input signal were recorded using a single-axis piezoelectric shear accelerometer (ACCELEROMETER, ICP® Model 352A24) with a sensitivity of $10.07\ mV/(m/s^2)$ (or $98.7mV/g$).

The input signal followed a linear time/frequency relationship, with the frequency increasing at a rate of approximately $21.35\ Hz$ per second. This means that as time progresses, the frequency of the input signal gradually and consistently increases.

By conducting this characterization process, we aimed to gain insights into the VH's response to different frequencies and ascertain its frequency-dependent behavior. This information is vital for understanding the VH's capabilities and limitations, ultimately contributing to the overall understanding of its performance during the subsequent experiments.

Initially, we conducted various measurements on the headphone's ear cup, encompassing both its external and internal components, which come into contact with the skin. Subsequently, we measured five distinct points situated between the mandible and the neck, as illustrated in Figure 4.1.

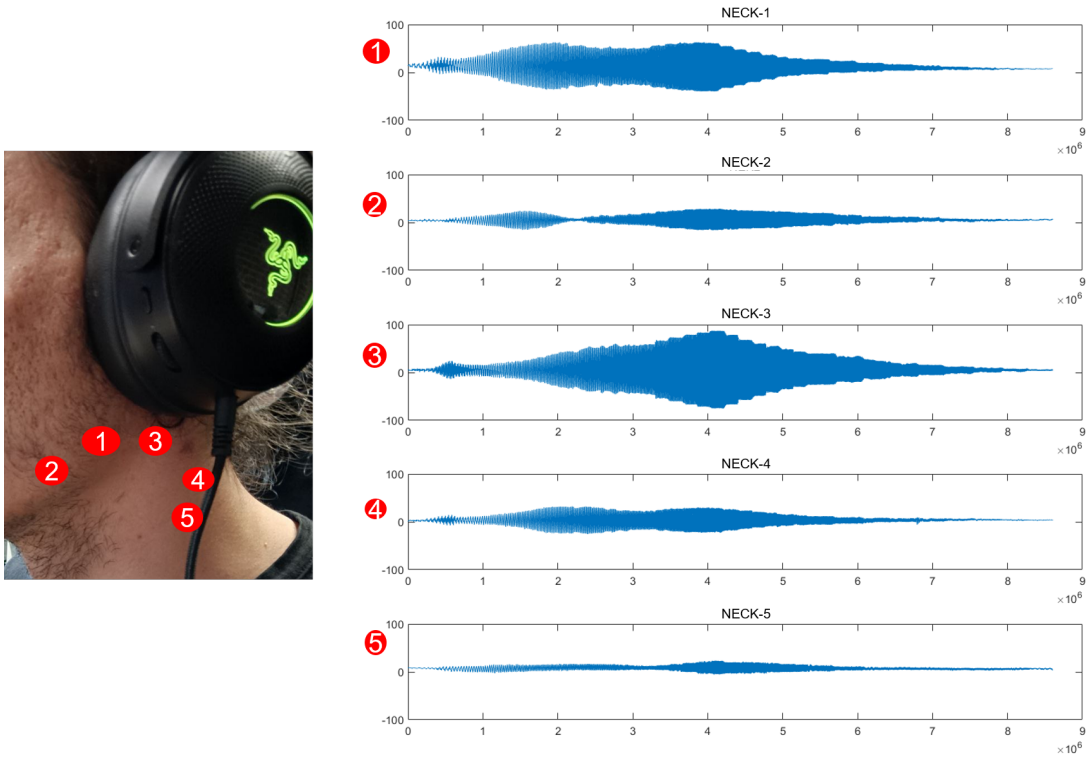

Figure 4.1: Left Illustration of the measuring points. **Right** Measured output of the accelerometer plotted as mV as a function of time (μs).

Examining the right side of Figure 4.1, it becomes evident that points 2, 3, and 4 exhibit significant signal attenuation due to the skin's dynamics. Conversely, points 1 and 3 display higher signal amplitudes, indicating lesser attenuation. This is partially attributed to their proximity to the ear cup. Consequently, our focus in Figure 4.1 shifted towards points 1 and 3, where we studied the extent of attenuation in comparison to a point located on the external ear cup (referred to as point EXT in Figure 4.2).

In order to assess the damping effect of vibrations and the amplitude loss in the input signal, we conducted measurements on both the VH (external part of the ear cup) and the neck. Fig. 4.2 illustrates the specific measuring points used, with one point on the VH itself, one on the mandible, and another on the neck. The recorded vibrations were converted from millivolts (mV) to gravitational units (m/s^2) using the conversion factor specified in the accelerometer's datasheet.

Analyzing the data from Fig. 4.2, we observed the presence of resonance peaks in the actuator's

frequency response, particularly at approximately $\sim 96Hz$. At this resonance peak, we observed a damping effect of 36.5% at point 2 (on the neck) and a larger damping effect of 60.21% at point 1 (on the mandible). Although this frequency corresponded to the strongest vibration within the measured range, it exhibited inefficiency from an energetic perspective and generated audible noise, which was not desirable for our experimental setup. Our goal was to ensure that participants could feel the vibrations without being distracted by audible noise.

Around the frequency of 60 Hz, we observed a lower damping effect of 13.85% at point 2 and 16.92% at point 1. The frequency range between 60 Hz and 72 Hz exhibited a more consistent amplitude response, making it a suitable choice for vibration stimuli. However, we needed to select a frequency that would not result in audible feedback from the vibrations. Frequencies below 60 Hz were avoided due to the rapid decrease in vibration strength even if the tendency of the audible noise resulting from the vibrations becomes imperceptible.

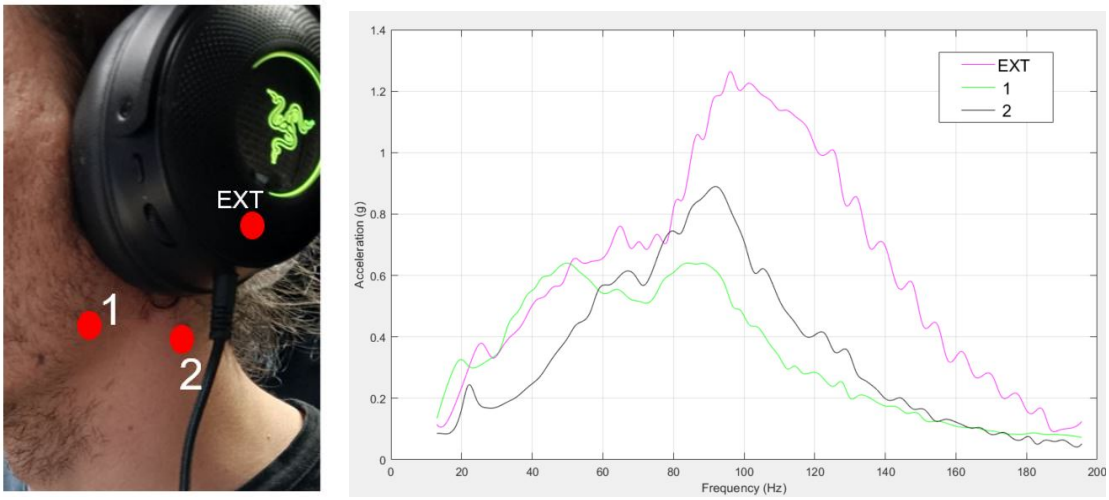

Figure 4.2: Left Illustration of the measuring points. One point is on the external part of the VH's cup. Two points are on the body: one on the mandible, one in the neck, about two cm right below the ear cup. **Right** Measured output of the accelerometer plotted as frequency vs acceleration (m/s^2).

By carefully considering these factors, we aimed to strike a balance between perceptible vibrations and minimizing any potential auditory feedback, ensuring an optimal experience for participants in our experiment. Therefore, we decided to work with a frequency of 60 Hz, as it provided easily perceptible

vibrations, even if still being slightly audible. However, the audible component will be masked using a pink noise during the experimental condition where only haptic feedback will be present. More details will be provided in the next section.

4.4 Stimuli and Setup

4.4.1 Setup

The delivery of haptic and auditory stimuli for this experiment relies on the utilization of the Razer© Kraken V3 HyperSense vibration headphones. The entire experiment is developed and implemented using the Unity3D game engine, specifically version 2020.3.20f1 [235]. To provide an immersive virtual experience, we employ the commercial Oculus Quest 2 headset [236]. Within the virtual scene, an object resembling a sphere is defined as both the haptic source and the audio source. This sphere object is positioned randomly in relation to the user's head while maintaining a consistent distance. The generation of audio and haptic stimuli is contingent upon the position of the sphere object and the specific experimental conditions being investigated.

The experiment comprises three distinct conditions: Haptic-only, audio-only, and audio-haptic. In each condition, the type of stimuli employed varies accordingly.

In the haptic-only condition, the Razer© Kraken V3 HyperSense headphones exclusively deliver haptic feedback to the participants based on the position of the sphere object relative to their head.

In the audio-only condition, participants solely receive auditory feedback via the Razer© Kraken V3 HyperSense headphones. The audio stimuli are generated based on the position of the sphere object within the virtual environment.

In the audio-haptic condition, participants experience a combination of both haptic and auditory feedback. The stimuli in this condition are determined by the position of the sphere object, and participants receive simultaneous audio and haptic cues through the Razer© Kraken V3 HyperSense headphones.

By delineating the specific stimuli employed in each experimental condition, we can effectively investigate the impact and effectiveness of haptic-only, audio-only, and audio-haptic feedback in the context of object localization within the VR environment.

4.4.2 Tactile Stimuli

The haptic stimuli in our experiment consist of a pulse signal with a frequency of 60 Hz and a duration of one second. The amplitude of the haptic signal is dynamically adjusted in real-time based on the orientation of the user's head in the virtual environment (VE). Among different possible functions for intensity tuning, we choose to use a method that mimics a simple HRTF for specialized audio in 3D space with the objective of making the haptic feedback easy to understand and use. Therefore, to determine the amplitude for each side of the VH (vibration headphone), we employ a generic HRTF as depicted below:

$$L = N * (1 + sin(\alpha))/2 \qquad 0 \leq \alpha < 360^\circ$$

$$R = N * (1 - sin(\alpha))/2 \qquad 0 \leq \alpha < 360^\circ$$

Here, N represents the number of vibration levels, and α denotes the angular position of the haptic source object relative to the user's head. The variables L and R indicate the vibration intensities applied to the left and right sides of the VH, respectively. As an example, when $\alpha = 0^\circ$ (indicating that the object is directly in front of the user), the vibration intensity on both sides of the VH is equal ($L = R = N/2$). As α increases (indicating that the object is rotating on the user's right side), the intensity on the right side of the VH increases and reaches its maximum at 90° ($R = N$), while the intensity on the left side decreases, reaching its minimum (i.e., $L = 0$ for $\alpha = 90^\circ$).

To provide a visual representation of these angular positions and the associated haptic stimuli, refer to Fig. 4.3. The figure illustrates the user's head at the center and the Haptic Source (the object with a vibration component) at five different angular positions. This visualization helps demonstrate how the amplitude of the vibrations varies based on the relative angular position of the haptic source object with respect to the user's head, thus illustrating the logic behind our HRTF.

To ensure that participants rely solely on their somatosensory ability during the task, we introduce a pink noise background audio in the haptic-only condition. By incorporating this auditory stimulus, we effectively mask any slightly audible cues that may arise from the haptic signal. This approach guarantees that participants focus exclusively on the tactile sensations provided by the haptic feedback, eliminating any potential interference from auditory cues.

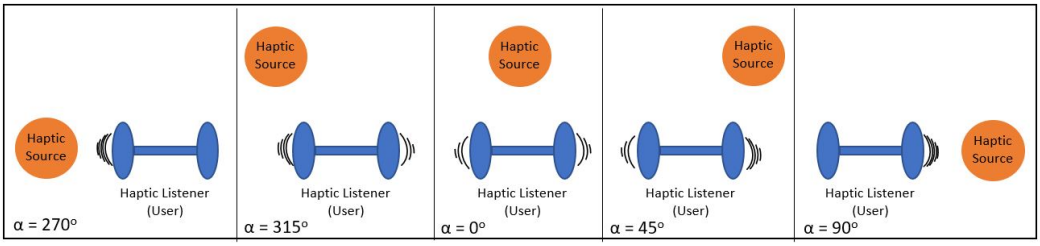

Figure 4.3: Representation of the VH and the different intensities for the left and right sides depending on the angular position of the haptic source. We have only one side vibrating in the first and last schema, both sides vibrating equally in the third schema ($\alpha = 0°$) and the intermediate level where one side is vibrating more than the other side ($\alpha = 315°$ and $\alpha = 45°$)

4.4.3 Auditory Stimuli

In both the audio-only and audio-haptic conditions, the auditory stimuli consist of pink noise. To implement the audio component, the sphere object, which serves as the haptic source, will be equipped with a Unity Audio Source component. To achieve a 3D sound simulation, we will utilize the Unity Audio Spatializer HRTF SDK [237], configuring the Spatial Blend parameter to 3D. This setting enables us to play the audio stimuli, creating the perception of sound originating from specific positions in virtual space.

The implementation of the 3D sound simulation involves reducing the audio to a single channel and then applying attenuation (altering the volume) based on the distance and position of the listener, which, in this case, is the user's head. As the distance between the audio source and the user remains fixed, the audio sound will only change in response to the angular position of the object relative to the user's head. Furthermore, there is no implementation of the Doppler effect of the audio source. The HRTF of the audio in the 3D space closely aligns with the haptic stimuli HRTF within the virtual environment, as depicted in Figure 4.3.

4.5 Experimental Protocol

4.5.1 Participants

We recruited a total of twenty-three participants. Two participants withdrew from the experiment due to motion sickness provoked by the Virtual Environment. Twenty-one healthy volunteers participated

in these experiments (15 males and 6 females, aged M = 26.4, 2 and SD = 5.23). Eighteen participants were right-handed, and three were left-handed. None of the participants declared problems of hearing loss or tactile numbness in the area surrounding the ears. Each participant took part in the experiments voluntarily and signed the informed consent before the start of the experiment, and the ethical committee of Lille University approved the experiment.

4.5.2 Procedure

In the experimental setup, participants were comfortably seated in chairs and equipped with the Meta Oculus 2 headset connected to a PC running the simulation. Before the experiment began, participants were given an explanation of how the stimuli behaved in relation to the position of a sphere. The initial trial was designed to familiarize participants with the haptic, audio, or audio-haptic feedback provided by the system. This familiarization phase aimed to ensure that participants had a clear understanding of the information conveyed by the stimuli in relation to the position of the sphere that was visible in this first trial. Participants moved their heads from side to side to become familiar with the stimuli, but no training was provided during this phase. We intentionally excluded preliminary training to assess participants' performance without any prior guidance. Once participants indicated that they were ready, the experimental session began.

The main task involved searching for an invisible sphere positioned randomly in the space around the participants while maintaining a fixed distance. To locate the sphere, participants were required to rotate their heads and rely on the provided stimuli, which could be haptic-only, audio-only, or a combination of both (audio-haptic). Once participants believed they had found the sphere, they would press a button on the Oculus controller. Subsequently, the sphere would become visible, allowing participants to verify its actual location. A brief pause of 3 seconds occurred before the start of the next trial, during which time the stimuli were turned off. This decision was made to prevent a decrease in participants' sensitivity to vibration. Previous research by Hochreiter et al. [238] indicates that prolonged exposure to vibrations can increase the tactile threshold in the hand, resulting in decreased tactile sensitivity. We believe this effect may also occur in the area surrounding the ears, and for this reason, we employed this design choice. Fig 4.4 illustrates the top view of the simulated 3D environment during a trial.

To ensure a balanced experimental design, the order of the three experimental conditions was counterbalanced among the participants. Each participant performed a total of 30 trials for each condition, resulting in a comprehensive dataset for analysis. To mitigate fatigue, a 3-minute break was provided to participants at the end of each experimental condition. Overall, the experiment lasted approximately 30 minutes for each participant, taking into account the time required for task completion and the scheduled pauses.

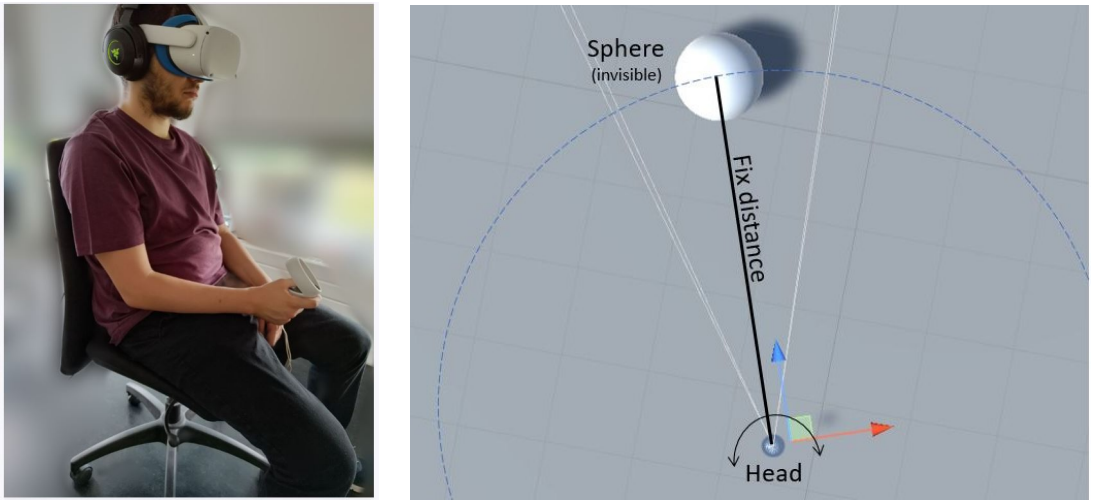

Figure 4.4: Left A participant during the experimental session. **Right** Experiment's top view during a trial. Participants rotated their heads and stopped in the direction of the sphere based on the provided feedback (audio-alone, haptic-alone, or audio-haptic). The invisible sphere (visible in the picture) remained at a fixed distance while the angular position was randomized for each trial. The white lines in the image represent the participants' field of view.

4.5.3 Data Analyses

In the course of the experiment, we meticulously gathered data on the angular position of the sphere and the corresponding angular error, which denotes the disparity between the sphere's position and the participant's head rotation. These measurements were utilized to ascertain the accuracy of participants across the various experimental conditions. Moreover, we took also the completion time for each condition and participant to evaluate possible differences.

Upon the conclusion of the experimental session, we proceeded to solicit participants' subjective evaluations of the different conditions, focusing on several key parameters. Participants were asked to

provide ratings based on the following criteria: Difficulty, Preference and Enjoyment. These ratings aimed to gauge participants' perceptions and subjective experiences within each condition.

Furthermore, we invited participants to provide additional comments pertaining to their overall experience and any specific features they wished to see implemented in relation to spatial haptic technology. This open-ended question allowed participants to offer their thoughts, insights, and suggestions, providing valuable qualitative feedback that complemented the quantitative data collected during the experiment.

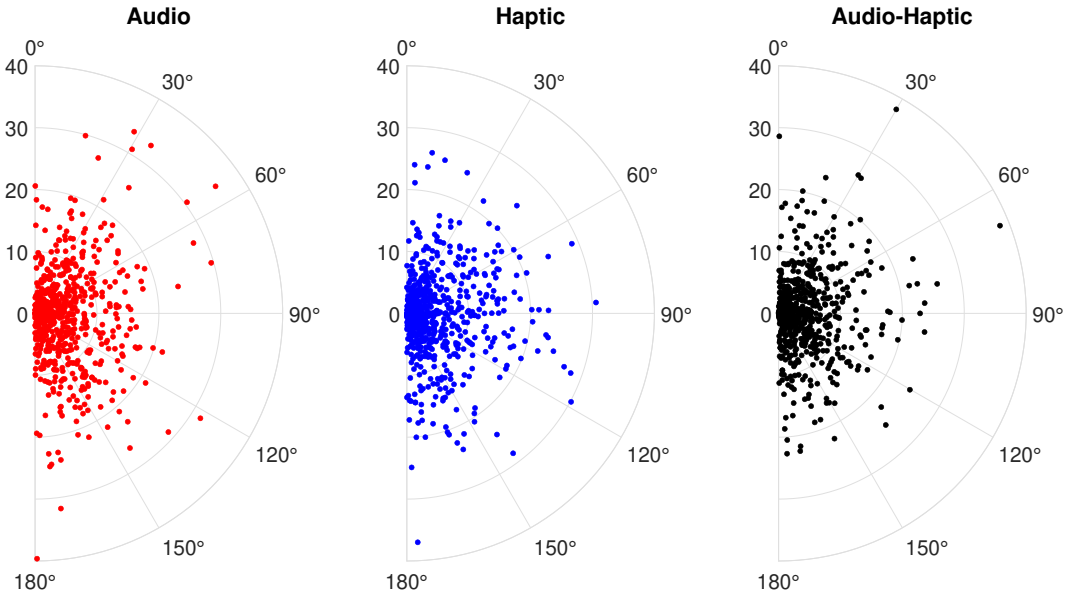

Figure 4.5: The polar plot illustrates the three experimental conditions for all participants. The angular axis represents the angular position of the sphere in each trial. Specifically, angles less than $90°$ correspond to objects located on the left side of the participant's head forward direction, while angles greater than $90°$ represent objects on the right side. The y-axis represents the angular error associated with the angular position of the sphere. It provides a measure of the deviation between the participants' intended direction and the actual position of the sphere.

4.6 Results

For each trial and condition across all subjects, we recorded the angular error, which represents the difference between the forward direction of the head and the angular position of the haptic source object, as well as the sphere's angular position. In total, we collected 1890 data points, consisting

of 30x3 angular error-position pairs. To visualize the results for all participants and conditions, we present a polar plot showing the distributions of sphere position and angular error in Fig. 4.5.

The distribution of trials between the left and right sides appears to be well-balanced for each participant. Specifically, in all experimental conditions, every participant encountered an almost equal number of spheres on their left and right sides (difference $< 5\%$).

The box plot displayed in Fig. 4.6 presents the angular error for each experimental condition across all participants. The median angular error was $6.547°$ for condition A, $5.910°$ for condition H, and $5.511°$ for condition AH. These results indicate that, overall, participants achieved lower errors in the AH condition, followed by the H condition and then the A condition.

To conduct a more detailed analysis of our participant population, we requested them to self-identify as either gamers or non-gamers. This information was gathered using the following definition: a gamer is someone who spends an average of at least 8 hours per week playing video games.
Interestingly, when dividing the participants into gamers (14/21) and non-gamers (7/21), a slightly different pattern emerges. In our results, gamers showed similar accuracy for conditions A and H ($6.394°$) but had lower errors in the AH condition ($5.445°$). In contrast, non-gamers demonstrated better accuracy in the H condition ($5.126°$), followed by the AH condition ($5.589°$), and finally, the A condition ($7.092°$).

4.6.1 Analysis

To analyze the effects of the three experimental conditions (A, H, AH) on accuracy, measured as angular error, we performed an analysis.
At first, we have checked for normality using the Kolmogorov-Smirnov test. The test shows that the data not follow a normal distribution. For this reason, we have used the Wilcoxon rank-sum test,a nonparametric alternative to the two- sample t-test which is based solely on the order in which the observations from the two samples fall. Performing this test, our results shows a statistical significant difference between Audio and Audio-Haptic conditions ($p = 0.0022$). No other significant difference was found.

Additionally, the data underwent further analysis by dividing participants into two groups: gamers and non-gamers. Statistically significant differences in the gamers' sub-group was found for Audio and

Audio-Haptic ($p = 0.0282$) and for Haptic and Audio-Haptic ($p = 0.0194$). No statistical difference was found for Audio and Haptic conditions. Results for non-gamers revealed significant differences in accuracy between the Audio and Audio-Haptic conditions ($p = 0.0107$) and between the Audio and Haptic conditions ($p = 0.0038$). However, no statistically significant difference was found between Haptic and Audio-Haptic conditions in the non-gamers group. Nevertheless, it's essential to note that the sample size of the non-gamers group was relatively small ($N = 7$). Therefore, the statistical analysis and comparison relative to gamers and not gamers should be interpreted with caution.

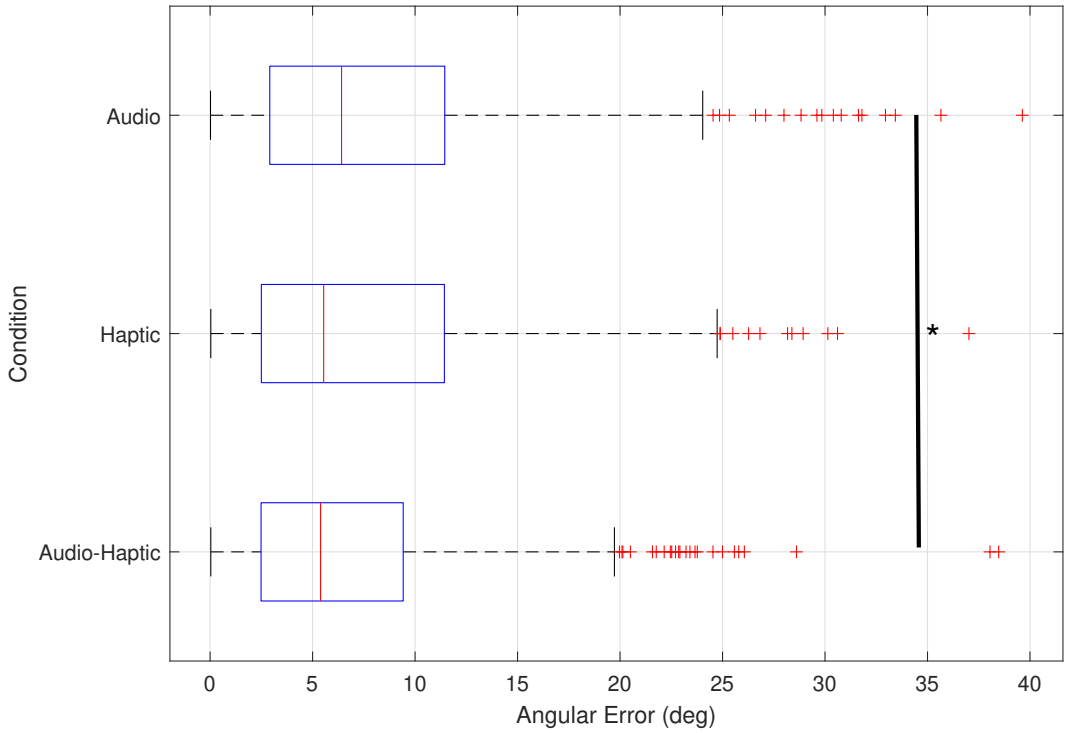

Figure 4.6: Boxplot visualization illustrating the distribution of data across the experimental conditions Audio(A), Haptic(H) and Audio-Haptic(AH). Each box represents the interquartile range (IQR) with the median indicated by the red vertical line. The whiskers extend to the most extreme non-outlier data points within 1.5 times the IQR, while any outliers are shown as individual points (red cross) beyond the whiskers. *$p < 0.05$,

4.6.2 Completion Time

In relation to completion time, we recorded the duration for each participant and condition. The completion times for the audio, haptic, and audio-haptic conditions are presented in the table below:

Table 4.1: Mean and standard deviation Completion time for our three conditions

	Completion Time		
	Session	Singular Trial	STD
Audio	5 min 54 s (354.3 s)	11.81 s	2.02
Haptic	6 min 20 s (380 s)	12.6 s	2.17
Audio-Haptic	6 min 14 s (374.2 s)	12.47 s	1.78

Overall, our participants took an additional 20 s to complete the audio-haptic condition and 26 s longer for the haptic-alone condition, both compared to the audio-alone condition. Performing a one-way ANOVA did not reveal any statistically significant difference in completion time among our experimental conditions. Therefore, the completion time was not influenced by the experimental conditions.

4.6.3 Subjective Evaluation

Following the experimental session, participants were asked to provide their subjective evaluations of the various experimental conditions, including Difficulty, Preference, and Enjoyment. These collected evaluations are visually represented in Fig. 4.7.

From the bar plot, it is evident that participants generally perceived the audio-alone condition as the most challenging (10/21), followed by the haptic condition (6/21), and then the audio-haptic condition (5/21). Regarding preference, the Haptic alone condition received the highest rating (14/21), followed by the audio-haptic condition (6/21), and lastly, the audio-alone condition (1/21). When it comes to enjoyment, the haptic alone condition ranked the highest (10/21), closely followed by the audio-haptic condition (9/21), while the audio-alone condition received the lowest rating

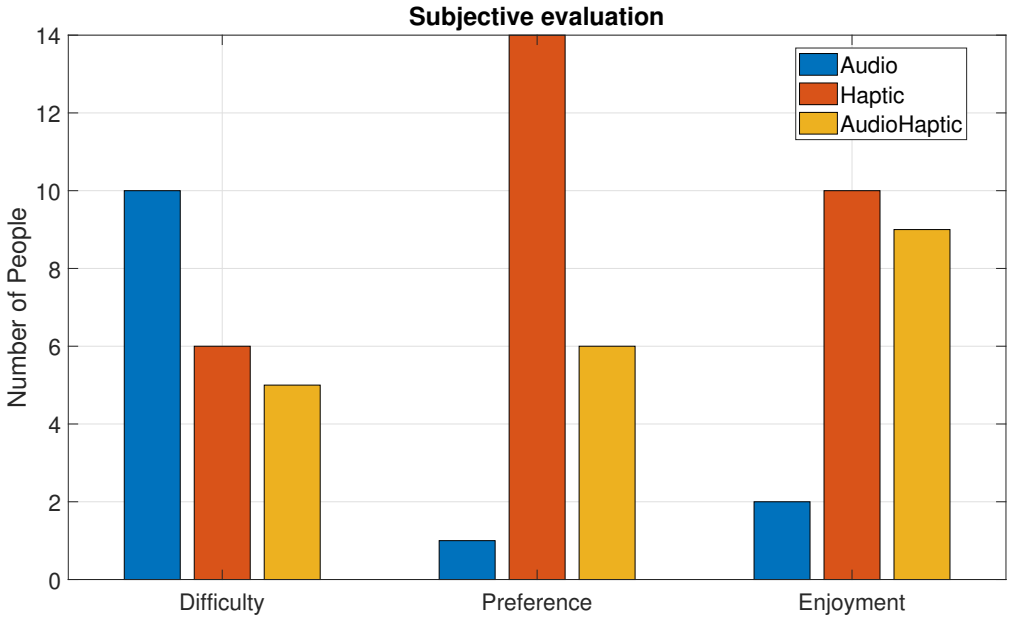

Figure 4.7: Participants' subjective evaluation on Difficulty, Preference and Enjoyment criteria for the experimental conditions.

(2/21).

4.7 Discussion

In this study, we evaluated the potential effectiveness of vibration headphones (VH) to convey directional information in an object localization task in a VR environment. Participants performed tasks with haptic-only, audio-only, and audio-haptic feedback conditions. At first, we performed a characterization of the Razer© Kraken V3 HyperSense headphones' mechanical behavior that revealed a resonance peak at approximately 96 Hz.

From our analysis, the audio-haptic condition shows a statistically significant difference in terms of accuracy compared to audio-only and haptic-only conditions. Even if, in general, our participants had lower errors in the haptic-only condition (5.91^o) compared to the audio-only condition 6.55^o), this difference was not statistically significant. However, our results show that the auditory channel can be substituted with the tactile channel when conveying directional information. Moreover, haptic can also be used to reinforce directional information provided by the auditory channel. Indeed, our

results suggest that combining haptic and auditory cues provides a more accurate (5.51^o)) sense of object localization, probably due to the reinforcement of directional information provided by the two sensory channels. Nevertheless, these results confirm the idea that multimodal integration leads to better performances [104].

On the subjective evaluation, participants perceived the haptic-only condition as the most enjoyable and preferred feedback method, indicating a preference for tactile sensations. Although the audio-only condition was perceived as more challenging, the addition of haptic feedback led to a more enjoyable experience for participants, likely due to increased immersion and engagement. The fact that the audio-only condition was rated as preferred by only one participant and by two for enjoyment may be due to the type of audio feedback we chose for the experiment (pink noise).

Interestingly, the completion times for the three conditions were relatively similar, with participants taking an average of approximately 12 seconds for each trial. While the audio-only condition had the shortest completion time, the slight differences among conditions do not appear to significantly impact task performance.

Upon analyzing participants' additional comments, we observed that they found the method of providing information regarding the object position through haptic feedback to be easily understandable and user-friendly. Some participants used the terms 'easy adaptation' or 'intuitive' when referring to haptic feedback. The feeling was 'like someone was guiding me to find the object' a participant stated while an other says that 'haptic feedback was like someone was pushing me in the object direction'. Some participants talk about a 'barrier showing them where the object was'. Interestingly, six participants expressed that they were using audio at first to find an approximation of the object position then focusing more on haptic to 'be more precise', when they were under the audio-haptic condition.

Overall, our haptic-based HRTF method for conveying directional information for object localization in VR was appreciated by the participants, who showed and stated positive feedback during the open-ended question at the end of the experiential session.

4.8 Conclusions

With this experiment, we show the potential of VH for conveying directional information through the tactile sensory modality for object localization in VR scenarios. Our findings highlight that our haptic feedback was easy to use without the requirement of any training. Furthermore, the implemented HRTF algorithm, which for certain aspects mimics an audio-based HRTF for sound in 3D, was effective and did not bring poorer performances in terms of accuracy, showing, on the contrary, an improvement compared to audio alone. Moreover, our participant rated the haptic-only as the preferred one among our experiential conditions, highlighting the effectiveness of our method.

The integration of haptic feedback with VR holds exciting possibilities for immersive and engaging virtual experiences, with potential applications in gaming, navigation scenarios and beyond. Interestingly, our haptic-based method shows potential for sensory substitution, relative to directional information, in scenarios where the audio channel is busy with another task or not usable due to environmental noise. Moreover, our haptic-based method also opens the possibility for sensory reinforcement in navigation scenarios or in general in a situation where directional information is crucial and has to be well understood. Future work may explore variations in VH configurations and their impact on user experience, ultimately enhancing the design and implementation of haptic feedback systems in VR.

4.9 General Guidelines

This section summarizes the main finding drawn from chapter 4, and it aims to provide practical guidelines for designers and developers.

Practical Guidelines

- Vibrating headphones are effective in providing directional information in VR with a sensitivity of 7^o.
- Providing directional information with audio-alone has a sensitivity of 8^o.
- Providing directional information with the combination of audio-tactile pair improves sensitivity (6^o)

CHAPTER 5

CONCLUSIONS AND PERSPECTIVES

5.1 Conclusions

The present PhD thesis centers on sensory and cross-modal interactions, emphasizing the intricate dynamics of sensory synchronization and perception through the incorporation of haptic feedback into a multimodal system. It aims to provide valuable guidelines for designers and developers in this field.

Initially, the thesis examines the characterization of human sensing, focusing on vision, hearing, and the mechanisms underlying tactile encoding. It provides an overview of the current knowledge regarding the anatomical and mechanical properties of the skin and the organization and functionality of mechanoreceptors, with particular attention to the distribution of their densities throughout the body. The thesis also explores haptic perception, with an emphasis on topics such as two-point discrimination, vibration perception, and texture perception, especially in relation to the finger pad. Additionally, the work conducts an in-depth analysis of the extensive and diverse range of haptic technologies, with a special focus on technologies of interest, such as surface haptic. It defines multimodality, with a particular interest in multimodal integration, especially in the context of visuo-haptic and audio-haptic interactions. Furthermore, the thesis inspects the concept of haptic illusions and pseudo-haptic effects as part of this comprehensive exploration.

The relevance of audio-tactile sensory synchronization has been established as a critical factor influencing the perception of multimodal experiences. Sensitivity to this synchronization has been experimentally observed, particularly during active touch involving sliding gestures, where it was found to be less pronounced compared to experiments involving passive touch. Additionally, the presence of a visual cue alongside tactile stimuli does not consistently impact the detection of asynchronies, highlighting that a co-localized visual-tactile multimodal system does not necessarily improve sensitivity in such a scenario. Notably, our data revealed a dissociation between finger sliding velocity and the perception of sensory synchronization, indicating that finger velocity does not significantly influence the threshold estimation. However, self-generated movements inherent in active touch during sliding gestures can lead to the chronostasis phenomenon, affecting the estimation and, therefore, increasing the acceptable time delay between the stimuli.

When examining scenarios involving audio-tactile clicks, our study observed that our participants exhibited higher sensitivity to asynchronies when audio stimuli preceded the haptic modality with a delay, suggesting that this particular sequence should be avoided. Conversely, when audio feedback followed haptic with a delay below $109\ ms$, the two stimuli were perceived as synchronous, indicating that $109\ ms$ serves as a critical boundary not to be exceeded. Going deeper in our investigation, in the comparison between sighted and blind participants, an intriguing contrast emerged. Adhering to delay limits suitable for sighted individuals proved sufficient for both populations, as blind individuals demonstrated a greater tolerance for delays. This observation underscores the significance of inclusivity in hardware development, particularly concerning virtual buttons.

In the realm of visual-tactile interaction, the impact of 3D depth cues in an augmented reality (AR) interface for texture perception has been demonstrated to be significant. Introducing depth through stereoscopy and texture deformation reveals that our tactile perception is influenced in the case of smooth textures, while visual perception is affected by tactile feedback. Overall, we find that in this context, our tactile modality plays a prominent role in shaping the overall judgment of roughness. However, by adding stereoscopy and tuning texture deformation depth, we showed that it is possible to manipulate the final percept of smooth textures by making them rougher or smoother, an increase of 20% compared to the standard. These findings hold valuable insights for designers aiming to use stereoscopy and texture deformation to manipulate the final perception of a multimodal

experience where textures are employed.

By exploring the potential of tactile feedback in different fields, we employed vibration headphones to convey directional information within a virtual environment. Our system proved to be user-friendly, requiring no prior training while using a generic head-related transfer function. Indeed, our system exhibited a spatial sensitivity of approximately 7^o when relying solely on haptic feedback and approximately 8^o when using only auditory cues. Intriguingly, when evaluating audio-tactile reinforcing stimuli for directional information, our participants demonstrated even greater spatial sensitivity, approximately 6^o. This underscores the importance of conveying information multimodally and highlights the synergistic effect of integrating bounded information from different senses, ultimately leading to improved performance by reinforcing the conveyed information. This result is important as it confirms the Bayesian model of multisensory integration.

In summary, this work provides a deeper understanding of sensory synchronization and interaction with a focus on haptic feedback and provides useful guidelines for designers and the development of the future generation of multimodal haptic interfaces.

5.2 Perspectives

Building upon the insights gained from this thesis, there are several intriguing avenues for future research and development in the field of sensory and cross-modal interaction. These perspectives highlight areas where innovative solutions can be devised and novel contributions can be made.

Future research holds the potential to delve deeper into the intricacies of refining sensory synchronization mechanisms, particularly by exploring innovative methods to minimize asynchronies between audio and tactile modalities. However, as elaborated in Chapter 2 through the literature, sensory synchronization is intricately tied to the specific task at hand, to the user's attention focus or the device in use (as we saw in section 2.5) and all of this complicates the establishment of universal guidelines for designing generic multimodal systems. Therefore, gaining a more comprehensive understanding of the underlying mechanisms of sensory synchronization becomes pivotal. This may entail conducting in-depth investigations into the neuroscience aspects, with a particular emphasis on identifying the brain activation zones associated with sensory synchronization. Such insights could pave the way for

more nuanced and effective approaches to optimizing multimodal user experiences.

Investigating haptic illusions and pseudo-haptic effects offers exciting opportunities for researchers as they can dig deeper into understanding how these phenomena can be harnessed to create unique and immersive experiences in fields such as virtual environments. Moreover, to further improve inclusivity, designers and developers can focus on creating haptic interfaces that adapt also to population with varying degrees of sensory perception such as visual impaired people. This includes those with visual or auditory impairments, where haptic feedback can play a crucial role in conveying information. Indeed, developing systems that adapt to users' preferences and sensitivities can enhance the overall user experience in fields such as virtual and augmented reality. Furthermore, developing user-friendly haptic interfaces, as demonstrated in this thesis in chapter 4, remains a key priority, and future systems should be designed for easy adoption and accessibility without the need for extensive training. Moreover, multimodal haptic interfaces can be used to simulate physical interactions, enhancing learning experiences in areas such as science and engineering. Indeed, researchers and developers can explore applications in medical training, physical therapy, and remote patient consultations with more integration of haptic feedback, as this can positively benefit and improve the delivery of healthcare services as well as education.

The exploration of the interplay between 3D depth cues and haptic feedback, as discussed in Chapter 3, continues to be a captivating field of study. In future research, there is potential to integrate stereoscopic vision through VR headsets, augmented by head-tracking and hand-tracking systems to facilitate visual-tactile interactions. This investigation is crucial because the use of a virtual hand, as opposed to a physical one, may introduce variations in both visual and tactile perception. Another idea could be the inclusion of auditory feedback in our existing system (section 3.1) opens doors to examining its impact on visual and tactile perception, a domain that remains largely unexplored. As research progresses, there is an opportunity for further validation and enhancement of multisensory models, such as the Bayesian model of multisensory integration, which can offer a deeper comprehension of how diverse sensory modalities interact and collectively shape perception.

In various contexts, particularly in navigation, the utilization of vibration headphones holds significant potential. As elaborated in Chapter 4, our system showcased remarkable precision in localizing

virtual objects, underscoring its versatility beyond the realm of VR. Indeed, there is room for further evolution and adaptation, positioning it as a valuable navigation tool in the physical world. By seamlessly integrating our HRTF with widely-used applications like Google Maps, both sighted and visually impaired individuals can benefit from real-time directional guidance during their journeys. This innovative approach empowers users to align themselves correctly while walking, freeing the auditory channel for other essential information, such as alerts for obstacles or potentially hazardous environments, among others.

In conclusion, the research presented in this thesis opens doors to a multitude of possibilities in the realm of sensory synchronization and cross-modal interactions, particularly focusing on haptic feedback. As designers and developers continue to explore these perspectives, they have the potential to shape the future of multimodal haptic interfaces and advance our understanding of sensory perception in multiple fields such as virtual and augmented reality environments.

BIBLIOGRAPHY

[1] P. Barthelmess and S. Oviatt, "Multimodal interfaces: combining interfaces to accomplish a single task," *HCI Beyond the GUI, Morgan Kauffman*, 2008.

[2] "The multitouch's project website, https://multitouch-itn.eu/," 2021. Accessed: 2023-08-02.

[3] F. Quek, D. McNeill, R. Bryll, S. Duncan, X.-F. Ma, C. Kirbas, K. E. McCullough, and R. Ansari, "Multimodal human discourse: Gesture and speech," *ACM Trans. Comput.-Hum. Interact.*, vol. 9, p. 171–193, Sept. 2002.

[4] E. B. Goldstein and L. Cacciamani, *Sensation and perception*. Cengage Learning, 2021.

[5] E. Vezzoli, *Tactile feedback devices: friction control and texture generation*. PhD thesis, University of Lille, 2016.

[6] C. Zimmer, "Our strange, important, subconscious light detectors," *Discover Magazine. Retrieved*, pp. 02–18, 2012.

[7] "Optography, anatomy of the human eye. https://optography.org/catoptric-images/," 2021. Accessed: 2023-10-23.

[8] D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *The Journal of Physiology*, vol. 160, no. 1, pp. 106–154, 1962.

[9] S. Winkler and D. Min, "Stereoscopic image quality compendium," *ICICS 2011 - 8th International Conference on Information, Communications and Signal Processing*, 12 2011.

[10] "Auditory sustem schematic, thompsons road physiotherapy, https://www.britannica.com/science/inner-ear," 2021. Accessed: 2023-10-23.

[11] S. Van der Jeught, J. Aerts, A. Bradu, A. Podoleanu, and J. Buytaert, "Full-field thickness distribution of human tympanic membrane obtained with optical coherence tomography," *Journal of the Association for Research in Otolaryngology : JARO*, vol. 14, 05 2013.

[12] "Distribution of frequencies along the basilar membrane of the cochlea. *Encyclopædia Britannica*, https://www.britannica.com/science/inner-ear," 2021. Accessed: 2023-10-23.

[13] J. Casale, P. F. Kandle, I. Murray, and N. Murr, "Physiology, cochlear function," 2018.

[14] N. Martín-Alguacil, D. Ignacio de Gaspar, J. Schober, and D. Pfaff, "Somatosensation: End organs for tactile sensation," *Neuroscience in the 21st Century. New York, NY: Springer. http://dx. doi. org/10.1007/978-1-4614-1997-6_27*, 2013.

[15] V. Hayward, "A brief overview of the human somatosensory system," *Musical haptics*, pp. 29–48, 2018.

[16] W. Maurel, D. Thalmann, Y. Wu, and N. M. Thalmann, *Biomechanical models for soft tissue simulation*, vol. 48. Springer, 1998.

[17] A. Pérez-Sánchez, E. Barrajón-Catalán, M. Herranz-López, and V. Micol, "Nutraceuticals for skin care: A comprehensive review of human clinical studies," *Nutrients*, vol. 10, no. 4, p. 403, 2018.

[18] R. W. Van Boven and K. O. Johnson, "The limit of tactile spatial resolution in humans: grating orientation discrimination at the lip, tongue, and finger," *Neurology*, vol. 44, no. 12, pp. 2361–2361, 1994.

[19] M. Wiertlewski and V. Hayward, "Mechanical behavior of the fingertip in the range of frequencies and displacements relevant to touch," *Journal of biomechanics*, vol. 45, no. 11, pp. 1869–1874, 2012.

[20] S. Derler and G.-M. Rotaru, "Stick–slip phenomena in the friction of human skin," *Wear*, vol. 301, no. 1-2, pp. 324–329, 2013.

[21] H. C. Bastian, "The "muscular sense"; its nature and cortical localisation," *Brain*, vol. 10, no. 1, pp. 1–89, 1887.

[22] U. Proske and S. C. Gandevia, "Kinesthetic senses," *Comprehensive Physiology*, vol. 8, no. 3, pp. 1157–1183, 2011.

[23] D. A. T. Guzman, *Generation and control of tactile feedback with longitudinal ultrasonic vibration and human-in-the-Loop analysis*. PhD thesis, Université de Lille, 2021.

[24] D. Purves, G. J. Augustine, D. Fitzpatrick, L. C. Katz, A.-S. LaMantia, J. O. McNamara, S. M. Williams, *et al.*, "Mechanoreceptors specialized to receive tactile information," in *Neuroscience*, vol. 9, Sinauer Associates Sunderland, MA, 2001.

[25] H. Haeberle and E. A. Lumpkin, "Merkel cells in somatosensation," *Chemosensory perception*, vol. 1, pp. 110–118, 2008.

[26] S. McGurk, "Ganong's review of medical physiology–," *Nursing Standard*, vol. 24, no. 20, pp. 30–31, 2010.

[27] M. Nolano, V. Provitera, C. Crisci, A. Stancanelli, G. Wendelschafer-Crabb, W. R. Kennedy, and L. Santoro, "Quantification of myelinated endings and mechanoreceptors in human digital skin," *Annals of neurology*, vol. 54, no. 2, pp. 197–205, 2003.

[28] G. Corniani and H. P. Saal, "Tactile innervation densities across the whole body," *Journal of Neurophysiology*, vol. 124, no. 4, pp. 1229–1240, 2020.

[29] I. Birznieks, V. G. Macefield, G. Westling, and R. S. Johansson, "Slowly adapting mechanoreceptors in the borders of the human fingernail encode fingertip forces," *Journal of Neuroscience*, vol. 29, no. 29, pp. 9370–9379, 2009.

[30] C. Chapman, *Active touch*, pp. 35–41. Springer Berlin Heidelberg, 2009.

[31] dundeemedstudentnotes, "2-point discrimination; vibration and temperature sensation, https://dundeemedstudentnotes.wordpress.com/2012/04/12/2-point-discrimination-vibration-and-temperature-sensation/," Apr 2012. Accessed: 2023-08-02.

[32] R. H. LaMotte and M. A. Srinivasan, "Surface microgeometry: Tactile perception and neural encoding," in *Information Processing in the Somatosensory System: Proceedings of an International Symposium at the Wenner-Gren Center, Stockholm, 3—5 July, 1989*, pp. 49–58, Springer, 1991.

[33] A. Brisben, S. Hsiao, and K. Johnson, "Detection of vibration transmitted through an object grasped in the hand," *Journal of neurophysiology*, vol. 81, no. 4, pp. 1548–1558, 1999.

[34] S. Bensmaia, "Texture from touch," *Scholarpedia*, vol. 4, no. 8, p. 7956, 2009.

[35] M. Hollins, S. Bensmaïa, K. Karlof, and F. Young, "Individual differences in perceptual space for tactile textures: Evidence from multidimensional scaling," *Perception & Psychophysics*, vol. 62, pp. 1534–1544, 2000.

[36] T. Callier, H. P. Saal, E. C. Davis-Berg, and S. J. Bensmaia, "Kinematics of unconstrained tactile texture exploration," *Journal of neurophysiology*, vol. 113, no. 7, pp. 3013–3020, 2015.

[37] C. M. Greenspon, K. R. McLellan, J. D. Lieber, and S. J. Bensmaia, "Effect of scanning speed on texture-elicited vibrations," *Journal of The Royal Society Interface*, vol. 17, no. 167, p. 20190892, 2020.

[38] S. J. Lederman, "Tactile roughness of grooved surfaces: The touching process and effects of macro-and microsurface structure," *Perception & Psychophysics*, vol. 16, no. 2, pp. 385–395, 1974.

[39] A. I. Weber, H. P. Saal, J. D. Lieber, J.-W. Cheng, L. R. Manfredi, J. F. Dammann III, and S. J. Bensmaia, "Spatial and temporal codes mediate the tactile perception of natural textures," *Proceedings of the National Academy of Sciences*, vol. 110, no. 42, pp. 17107–17112, 2013.

[40] C. Bernard, *Perception of audio-haptic textures for new touchscreen interactions*. PhD thesis, Aix-Marseile Université, 2022.

[41] F. Martinot, "The influence of surface commensurability on roughness perception with a bare finger," in *proc. of Eurohaptics*, Citeseer, 2006.

[42] K. Kangur, M. Toth, J. Harris, and C. Hesse, "Everyday haptic experiences influence visual perception of material roughness," *Journal of Vision*, vol. 19, p. 300a, 09 2019.

[43] H. Culbertson, S. B. Schorr, and A. M. Okamura, "Haptics: The present and future of artificial touch sensation," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 1, no. 1, pp. 385–409, 2018.

[44] D. WANG, Y. GUO, S. LIU, Y. ZHANG, W. XU, and J. XIAO, "Haptic display for virtual reality: progress and challenges," *Virtual Reality & Intelligent Hardware*, vol. 1, no. 2, pp. 136–162, 2019.

[45] A. Kaci, *Méthodologie de commande de vibrations multimodales par modulation-démodulation synchrone : application au retour tactile "multi-touch*. PhD thesis, University of Lille, 2020.

[46] C. Basdogan, F. Giraud, V. Levesque, and S. Choi, "A review of surface haptics: Enabling tactile effects on touch surfaces," *IEEE Transactions on Haptics*, vol. 13, no. 3, pp. 450–470, 2020.

[47] "Voxel-man, https://www.voxel-man.com/simulators/dental/." Accessed: 2023-08-02.

[48] "Touch-x, https://www.3dsystems.com/haptics-devices/touch-x," Feb 2023. Accessed: 2023-08-02.

[49] Quanser, "Telepresence system, https://www.quanser.com/products/telepresence/," Apr 2021. Accessed: 2023-08-02.

[50] "Spidar-g2, http://www.nihonbinary.co.jp/products/vr/haptic/spidar/index_en.html." Accessed: 2023-08-02.

[51] "Gotouchvr, www.gotouchvr.com." Accessed: 2023-08-02.

[52] "European space agency, https://www.esa.int/enabling_support/space_engineering_technology/automation_and_robotics/haptic_devices." Accessed: 2023-08-02.

[53] "Prime x, https://www.manus-meta.com/products/prime-x-haptic." Accessed: 2023-08-02.

[54] "Haptx, https://haptx.com/." Accessed: 2023-08-02.

[55] Y. Kurita, "Chapter 1.3 - wearable haptics," in *Wearable Sensors* (E. Sazonov and M. R. Neuman, eds.), pp. 45–63, Oxford: Academic Press, 2014.

[56] L. Post, I. Zompa, and C. Chapman, "Perception of vibrotactile stimuli during motor activity in human subjects," *Experimental brain research*, vol. 100, pp. 107–120, 1994.

[57] T. A. Kern, C. Hatzfeld, and A. Abbasimoshaei, *Engineering haptic devices*. Springer, 2023.

[58] A. El Saddik, M. Orozco, M. Eid, and J. Cha, *Haptics technologies: Bringing touch to multimedia*. Springer Science & Business Media, 2011.

[59] "A Survey of Mid-Air Ultrasound Haptics and Its Applications," *IEEE Transactions on Haptics*, vol. 1412, no. c, pp. 1–1, 2020.

[60] T. Iwamoto, M. Tatezono, and H. Shinoda, "Non-contact method for producing tactile sensation using airborne ultrasound," in *Haptics: Perception, Devices and Scenarios: 6th International Conference, EuroHaptics 2008 Madrid, Spain, June 10-13, 2008 Proceedings 6*, pp. 504–513, Springer, 2008.

[61] "Ultraleap, https://www.ultraleap.com." Accessed: 2023-08-02.

[62] T. Romanus, S. Frish, M. Maksymenko, W. Frier, L. Corenthy, and O. Georgiou, "Mid-Air Haptic Bio-Holograms in Mixed Reality," pp. 2–6, 2019.

[63] "Touch hologram in mid-air," *ACM SIGGRAPH 2017 Emerging Technologies, SIGGRAPH 2017*, pp. 2–3, 2017.

[64] H. Kajimoto, "Enlarged electro-tactile display with repeated structure," in *2011 IEEE World Haptics Conference*, pp. 575–579, IEEE, 2011.

[65] C. Harrison and S. E. Hudson, "Providing dynamically changeable physical buttons on a visual display," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 299–308, 2009.

[66] Y. Ujitoko, S. Sakurai, and K. Hirota, "Influence of sparse contact point and finger penetration in object on shape recognition," *IEEE Transactions on Haptics*, vol. 13, no. 2, pp. 425–435, 2020.

[67] X. Ji, X. Liu, V. Cacucciolo, Y. Civet, A. El Haitami, S. Cantin, Y. Perriard, and H. Shea, "Untethered feel-through haptics using 18-μm thick dielectric elastomer actuators," *Advanced Functional Materials*, vol. 31, no. 39, p. 2006639, 2021.

[68] C. Hudin, J. Lozada, and V. Hayward, "Localized tactile stimulation by time-reversal of flexural waves: Case study with a thin sheet of glass," in *2013 World Haptics Conference (WHC)*, pp. 67–72, IEEE, 2013.

[69] T. Nara, M. Takasaki, S. Tachi, and T. Higuchi, "An application of saw to a tactile display in virtual reality," in *2000 IEEE Ultrasonics Symposium. Proceedings. An International Symposium (Cat. No. 00CH37121)*, vol. 1, pp. 1–4, IEEE, 2000.

[70] E. Vezzoli, T. Sednaoui, M. Amberg, F. Giraud, and B. Lemaire-Semail, "Texture rendering strategies with a high fidelity - capacitive visual-haptic friction control device," in *EuroHaptics*, 2016.

[71] E. Mallinckrodt, A. Hughes, and W. Sleator Jr, "Perception by the skin of electrically induced vibrations," *Science*, vol. 118, no. 3062, pp. 277–278, 1953.

[72] S. Grimnes, "Electrovibration, cutaneous sensation of microampere current," *Acta Physiologica Scandinavica*, vol. 118, no. 1, pp. 19–25, 1983.

[73] "Tanvas, https://tanvas.co/." Accessed: 2023-08-02.

[74] O. Bau, I. Poupyrev, A. Israr, and C. Harrison, "Teslatouch: electrovibration for touch surfaces," in *Proceedings of the 23nd annual ACM symposium on User interface software and technology*, pp. 283–292, 2010.

[75] S.-C. Kim, A. Israr, and I. Poupyrev, "Tactile rendering of 3d features on touch surfaces," in *Proceedings of the 26th annual ACM symposium on User interface software and technology*, pp. 531–538, 2013.

[76] T. Watanabe and S. Fukui, "A method for controlling tactile sensation of surface roughness using ultrasonic vibration," in *Proceedings of 1995 IEEE International Conference on Robotics and Automation*, vol. 1, pp. 1134–1139 vol.1, 1995.

[77] M. Amberg, F. Giraud, B. Semail, P. Olivo, G. Casiez, and N. Roussel, "Stimtac: A tactile input device with programmable friction," in *Proceedings of the 24th Annual ACM Symposium Adjunct on User Interface Software and Technology*, UIST '11 Adjunct, (New York, NY, USA), p. 7–8, Association for Computing Machinery, 2011.

[78] L. Winfield, J. Glassmire, J. E. Colgate, and M. Peshkin, "T-pad: Tactile pattern display through variable friction reduction," in *Second Joint EuroHaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (WHC'07)*, pp. 421–426, 2007.

[79] M. Biet, *Conception et contrôle d'actionneurs électro-actifs dédiés à la stimulation tactile*. PhD thesis, University of Lille, 2007.

[80] F.Giraud, M. Amberg, B. Lemaire-Semail, and G. Casiez, "Design of a transparent tactile stimulator," in *2012 IEEE Haptics Symposium (HAPTICS)*, pp. 485–489, 2012.

[81] D. Gueorguiev, E. Vezzoli, T. Sednaoui, L. Grisoni, and B. Lemaire-Semail, "The perception of ultrasonic square reductions of friction with variable sharpness and duration," *IEEE transactions on haptics*, vol. 12, no. 2, pp. 179–188, 2019.

[82] F. Giraud and C. Giraud-Audine, *Piezoelectric Actuators: Vector Control Method*. Elsevier, 2019.

[83] A. Minikes and I. Bucher, "Coupled dynamics of a squeeze-film levitated mass and a vibrating piezoelectric disc: numerical analysis and experimental study," *Journal of sound and vibration*, vol. 263, no. 2, pp. 241–268, 2003.

[84] A. Minikes and I. Bucher, "Comparing numerical and analytical solutions for squeeze-film levitation force," *Journal of fluids and structures*, vol. 22, no. 5, pp. 713–719, 2006.

[85] T. Sednaoui, E. Vezzoli, B. Dzidek, B. Lemaire-Semail, C. Chappaz, and M. Adams, "Experimental evaluation of friction reduction in ultrasonic devices," in *2015 IEEE World Haptics Conference (WHC)*, pp. 37–42, IEEE, 2015.

[86] N. Huloux, C. Bernard, and M. Wiertlewski, "Estimation of the modulation of friction from the mechanical impedance variations," *IEEE Transactions on Haptics*, vol. 14, no. 2, pp. 409–420, 2021.

[87] M. Wiertlewski, R. Fenton Friesen, and J. E. Colgate, "Partial squeeze film levitation modulates fingertip friction," *Proceedings of the national academy of sciences*, vol. 113, no. 33, pp. 9210–9215, 2016.

[88] M. Z. Baig and M. Kavakli, "Multimodal systems: Taxonomy, methods, and challenges," *CoRR*, vol. abs/2006.03813, 2020.

[89] S. Oviatt, "Advances in robust multimodal interface design," *Computer Graphics and Applications, IEEE*, vol. 23, pp. 62 – 68, 10 2003.

[90] S. Oviatt, "Multimodal interactive maps: designing for human performance," *Human-Computer Interaction*, vol. 12, no. 1-2, pp. 93–129, 1997.

[91] D. Bohus and E. Horvitz, "Facilitating multiparty dialog with gaze, gesture, and speech," in *International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction*, ICMI-MLMI '10, (New York, NY, USA), Association for Computing Machinery, 2010.

[92] M. J. Pitts, L. Skrypchuk, T. Wellings, A. Attridge, and M. A. Williams, "Evaluating user response to in-car haptic feedback touchscreens using the lane change test," *Adv. in Hum.-Comp. Int.*, vol. 2012, Jan. 2012.

[93] S. Oviatt and P. Cohen, "Perceptual user interfaces: Multimodal interfaces that process what comes naturally," *Commun. ACM*, vol. 43, p. 45–53, Mar. 2000.

[94] M. Turk, "Review article: Multimodal interaction: A review," *Pattern Recogn. Lett.*, vol. 36, p. 189–195, Jan. 2014.

[95] J. Zheng and W. Zhang, "Multimodal in-vehicle touch screens interactive system's design and evaluation," *Nunes I. (eds) Advances in Human Factors and Systems Interaction.*, vol. 1207, 2020.

[96] S. Oviatt, B. Schuller, P. Cohen, D. Sonntag, G. Potamianos, and A. Krüger, *The Handbook of Multimodal-Multisensor Interfaces: Foundations, User Modeling, and Common Modality Combinations - Volume 1*, vol. 14. Association for Computing Machinery and Morgan & Claypool, 2017.

[97] S. Sinclair and M. M. Wanderley, "A run-time programmable simulator to enable multi-modal interaction with rigid-body systems," *Interacting with Computers*, vol. 21, pp. 54–63, 11 2008.

[98] J. Leonard and J. Villeneuve, "Fast audio-haptic prototyping with mass-interaction physics," *HAID 2019 - International Workshop on Haptic and Audio Interaction Design*, 2019.

[99] B. E. Stein, T. R. Stanford, and B. A. Rowland, "The neural basis of multisensory integration in the midbrain: its organization and maturation," *Hearing research*, vol. 258, no. 1-2, pp. 4–15, 2009.

[100] L. P. G. Cortés, I. Q. de França, R. L. Gonçalves, and L. Pereira, "A design model roadmap for a multisensory experience," *DS 91: Proceedings of NordDesign 2018, Linköping, Sweden, 14th-17th August 2018*, 2018.

[101] D. Alais and D. Burr, "The ventriloquist effect results from near-optimal bimodal integration," *Current biology*, vol. 14, no. 3, pp. 257–262, 2004.

[102] H. McGurk and J. MacDonald, "Hearing lips and seeing voices," *Nature*, vol. 264, no. 5588, pp. 746–748, 1976.

[103] M. O. Ernst and H. H. Bülthoff, "Merging the senses into a robust percept," *Trends in cognitive sciences*, vol. 8, no. 4, pp. 162–169, 2004.

[104] M. O. Ernst and M. S. Banks, "Humans integrate visual and haptic information in a statistically optimal fashion," *Nature*, vol. 415, no. 6870, pp. 429–433, 2002.

[105] S. J. Lederman and S. G. Abbott, "Texture perception: studies of intersensory organization using a discrepancy paradigm, and visual versus tactual psychophysics.," *Journal of Experimental Psychology: Human perception and performance*, vol. 7, no. 4, p. 902, 1981.

[106] S. J. Lederman and L. A. Jones, "Tactile and haptic illusions," *IEEE Transactions on Haptics*, vol. 4, no. 4, pp. 273–294, 2011.

[107] A. Lécuyer, "Simulating haptic feedback using vision: A survey of research and applications of pseudo-haptic feedback," *Presence: Teleoperators and Virtual Environments*, vol. 18, no. 1, pp. 39–53, 2009.

[108] Y. Ujitoko and Y. Ban, "Survey of pseudo-haptics: Haptic feedback design and application proposals," *IEEE Transactions on Haptics*, vol. 14, no. 4, pp. 699–711, 2021.

[109] D. Goldreich and J. Tong, "Prediction, postdiction, and perceptual length contraction: a bayesian low-speed prior captures the cutaneous rabbit and related illusions," *Frontiers in psychology*, vol. 4, p. 221, 2013.

[110] S. K. Khuu, J. C. Kidd, and D. R. Badcock, "The influence of spatial orientation on the perceived path of visual saltatory motion," *Journal of Vision*, vol. 11, no. 9, pp. 5–5, 2011.

[111] J. Tong, V. Ngo, and D. Goldreich, "Tactile length contraction as bayesian inference," *Journal of neurophysiology*, vol. 116, no. 2, pp. 369–379, 2016.

[112] D. Goldreich, "A bayesian perceptual model replicates the cutaneous rabbit and other tactile spatiotemporal illusions," *PloS one*, vol. 2, no. 3, p. e333, 2007.

[113] M. Gori, G. Bertonati, C. Campus, and M. B. Amadeo, "Multisensory representations of space and time in sensory cortices," *Human Brain Mapping*, vol. 44, no. 2, pp. 656–667, 2023.

[114] G. Berkeley, A. Luce, and T. Jessop, *A new theory of vision*. Electric Book Company, 2001.

[115] M. S. Banks, R. N. Aslin, and R. D. Letson, "Sensitive period for the development of human binocular vision," *Science*, vol. 190, no. 4215, pp. 675–677, 1975.

[116] D. Burr and M. Gori, *Multisensory Integration Develops Late in Humans*, pp. 345–362. FRONTIERS IN NEUROSCIENCE, CRC Press, 2012.

[117] M. Gori, G. Sandini, and D. Burr, "Development of visuo-auditory integration in space and time," *Frontiers in integrative neuroscience*, vol. 6, p. 77, 2012.

[118] M. Gori, M. Del Viva, G. Sandini, and D. C. Burr, "Young children do not integrate visual and haptic form information," *Current Biology*, vol. 18, no. 9, pp. 694–698, 2008.

[119] R. L. Klatzky, S. J. Lederman, and V. A. Metzger, "Identifying objects by touch: An "expert system"," *Perception & psychophysics*, vol. 37, pp. 299–302, 1985.

[120] R. L. Klatzky and S. J. Lederman, "Identifying objects from a haptic glance," *Perception & Psychophysics*, vol. 57, pp. 1111–1123, 1995.

[121] R. L. Klatzky and S. J. Lederman, "Stages of manual exploration in haptic object identification," *Perception & psychophysics*, vol. 52, no. 6, pp. 661–670, 1992.

[122] M. Graf, "Categorization and object shape," *Towards a Theory of Thinking: Building Blocks for a Conceptual Framework*, pp. 73–101, 2010.

[123] T. Cooke, F. Jäkel, C. Wallraven, and H. H. Bülthoff, "Multimodal similarity and categorization of novel, three-dimensional objects," *Neuropsychologia*, vol. 45, no. 3, pp. 484–495, 2007.

[124] N. Gaissert, C. Wallraven, and H. H. Bülthoff, "Visual and haptic perceptual spaces show high similarity in humans," *Journal of vision*, vol. 10, no. 11, pp. 2–2, 2010.

[125] C. Wallraven, H. H. Bülthoff, S. Waterkamp, L. van Dam, and N. Gaißert, "The eyes grasp, the hands see: Metric category knowledge transfers between vision and touch," *Psychonomic bulletin & review*, vol. 21, pp. 976–985, 2014.

[126] M. A. Heller, "Visual and tactual texture perception: Intersensory cooperation," *Perception & psychophysics*, vol. 31, pp. 339–344, 1982.

[127] B. Barakat, A. R. Seitz, and L. Shams, "Visual rhythm perception improves through auditory but not visual training," *Current Biology*, vol. 25, no. 2, pp. R60–R61, 2015.

[128] D. Burr, M. S. Banks, and M. C. Morrone, "Auditory dominance over vision in the perception of interval duration," *Experimental Brain Research*, vol. 198, pp. 49–57, 2009.

[129] L. D. Rosenblum, J. W. Dias, and J. Dorsi, "The supramodal brain: implications for auditory perception," *Journal of Cognitive Psychology*, vol. 29, no. 1, pp. 65–87, 2017.

[130] M. B. Amadeo, C. Campus, and M. Gori, "Visual representations of time elicit early responses in human temporal cortex," *Neuroimage*, vol. 217, p. 116912, 2020.

[131] L. Shams, Y. Kamitani, and S. Shimojo, "What you see is what you hear," *Nature*, vol. 408, no. 6814, pp. 788–788, 2000.

[132] M. Gori, T. Vercillo, G. Sandini, and D. Burr, "Tactile feedback improves auditory spatial localization," *Frontiers in Psychology*, vol. 5, p. 1121, 2014.

[133] A. Tonelli, M. Gori, and L. Brayda, "The influence of tactile cognitive maps on auditory space perception in sighted persons," *Frontiers in psychology*, vol. 7, p. 1683, 2016.

[134] S. J. Lederman, "Auditory texture perception," *Perception*, vol. 8, no. 1, pp. 93–103, 1979.

[135] S. J. Lederman, R. L. Klatzky, T. Morgan, and C. Hamilton, "Integrating multimodal information about surface texture via a probe: relative contributions of haptic and touch-produced sound sources," in *Proceedings 10th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems. HAPTICS 2002*, pp. 97–104, IEEE, 2002.

[136] V. Jousmäki and R. Hari, "Parchment-skin illusion: sound-biased touch," *Current biology*, vol. 8, no. 6, pp. R190–R191, 1998.

[137] S. Guest, C. Catmur, D. Lloyd, and C. Spence, "Audiotactile interactions in roughness perception," *Experimental Brain Research*, vol. 146, pp. 161–171, 2002.

[138] S. Soto-Faraco and G. Deco, "Multisensory contributions to the perception of vibrotactile events," *Behavioural brain research*, vol. 196, no. 2, pp. 145–154, 2009.

[139] J. R. Timora and T. W. Budd, "Steady-State EEG and Psychophysical Measures of Multisensory Integration to Cross-Modally Synchronous and Asynchronous Acoustic and Vibrotactile Amplitude Modulation Rate," *Multisensory Research*, vol. 31, no. 5, pp. 391–418, 2018.

[140] D. Brahimaj, G. Esposito, A. Courtin, F. Giraud, B. Semail, O. Collignon, and A. Mouraux, "Temporal detection threshold of audio-tactile delays under conditions of active touch with and without a visual cue," in *IEEE World Haptics Conference (WHC), WHC 2023*, 2023.

[141] A. Kopinska and L. R. Harris, "Simultaneity constancy," *Perception*, vol. 33, no. 9, pp. 1049–1060, 2004.

[142] D. Brahimaj, G. Esposito, A. Courtin, F. Giraud, B. Semail, O. Collignon, and A. Mouraux, "Temporal detection threshold of audio-tactile delays under conditions of active touch with and without a visual cue," in *2023 IEEE World Haptics Conference (WHC)*, pp. 354–360, IEEE, 2023.

[143] D. Brahimaj, M. Ouari, A. Kaci, F. Giraud, C. Giraud-Audine, and B. Semail, "Temporal detection threshold of audio-tactile delays with virtual button," *IEEE Transactions on Haptics*, 2023.

[144] I. Birznieks, P. Jenmalm, A. W. Goodwin, and R. S. Johansson, "Encoding of direction of fingertip forces by human tactile afferents," *Journal of Neuroscience*, vol. 21, no. 20, pp. 8222–8237, 2001.

[145] N. P. Holmes and C. Spence, "Multisensory integration: Space, time and superadditivity," *Current Biology*, vol. 15, no. 18, pp. R762–R764, 2005.

[146] S. Soto-Faraco and G. Deco, "Multisensory contributions to the perception of vibrotactile events," *Behavioural Brain Research*, vol. 196, no. 2, pp. 145–154, 2009.

[147] I. J. Hirsh and C. E. Sherrick, "Perceived order in different sense modalities," *Journal of Experimental Psychology*, vol. 62, no. 5, pp. 423–432, 1961.

[148] W. Fujisaki and S. Nishida, "Audio-tactile superiority over visuo-tactile and audio-visual combinations in the temporal resolution of synchrony perception," *Experimental Brain Research*, vol. 198, no. 2-3, pp. 245–259, 2009.

[149] C. Spence, J. Driver, and J. C. Driver, *Crossmodal space and crossmodal attention*. Oxford University Press, 2004.

[150] V. Occelli, C. Spence, and M. Zampini, "Audiotactile temporal order judgments in sighted and blind individuals," *Neuropsychologia*, vol. 46, no. 11, pp. 2845–2850, 2008.

[151] N. Kitagawa, M. Zampini, and C. Spence, "Audiotactile interactions in near and far space," *Experimental brain research*, vol. 166, no. 3, pp. 528–537, 2005.

[152] M. Zampini, T. Brown, D. I. Shore, A. Maravita, B. Röder, and C. Spence, "Audiotactile temporal order judgments," *Acta psychologica*, vol. 118, no. 3, pp. 277–291, 2005.

[153] V. Occelli, C. Spence, and M. Zampini, "Audiotactile interactions in temporal perception," *Psychonomic Bulletin and Review*, vol. 18, no. 3, pp. 429–454, 2011.

[154] L. L. Kontsevich and C. W. Tyler, "Bayesian adaptive estimation of psychometric slope and threshold," *Vision Research*, vol. 39, no. 16, pp. 2729–2737, 1999.

[155] Stan Development Team, "Stan modeling language users guide and reference manual, version 2.17.0," 2017.

[156] Y. Rekik, E. Vezzoli, L. Grisoni, and F. Giraud, "Localized haptic texture: A rendering technique based on taxels for high density tactile feedback," *Conference on Human Factors in Computing Systems - Proceedings*, vol. 2017-May, pp. 5006–5015, 2017.

[157] E. Vezzoli, Z. Vidrih, V. Giamundo, B. Lemaire-Semail, F. Giraud, T. Rodic, D. Peric, and M. Adams, "Friction Reduction through Ultrasonic Vibration Part 1: Modelling Intermittent Contact," *IEEE Transactions on Haptics*, vol. 10, no. 2, pp. 196–207, 2017.

[158] T. Sednaoui, E. Vezzoli, B. Dzidek, B. Lemaire-Semail, C. Chappaz, and M. Adams, "Friction Reduction through Ultrasonic Vibration Part 2: Experimental Evaluation of Intermittent Contact and Squeeze Film Levitation," *IEEE Transactions on Haptics*, vol. 10, no. 2, pp. 208–216, 2017.

[159] G. Gescheider, *Psychophysics: The Fundamentals*. Taylor & Francis, 2013.

[160] W. Ben Messaoud, "Design and Control of a Tactile Stimulator for Real Texture Simulation : Application to Textile Fabrics Thèse Design and Control of a Tactile Stimulator for Real Texture Simulation : Application to Textile Fabrics," no. June, p. 147, 2016.

[161] A. Gelman and J. Hill, "Data analysis using regression and multilevel/hierarchical models," 2007. Includes bibliographical references (pages 575-600) and indexes.

[162] F. Kingdom and N. Prins, *Psychophysics: a practical introduction, Second Edition*. Academic Press, 2016.

[163] Stan Development Team, "shinystan: Interactive visual and numerical diagnostics and posterior analysis for bayesian models.," 2017. R package version 2.4.0.

[164] A. Moungou, *EEG frequency tagging to explore the cortical activity induced by the dynamic tactile exploration of textures*. PhD thesis, UCL-Université Catholique de Louvain, 2016.

[165] R. Kock, K. Gladhill, M. Ali, W. Joiner, and M. Wiener, "How movements shape the perception of time," *Trends in Cognitive Sciences*, vol. 25, 09 2021.

[166] M. K. Saleem, C. Yilmaz, and C. Basdogan, "Tactile Perception of Virtual Edges and Gratings Displayed by Friction Modulation via Ultrasonic Actuation," *IEEE Transactions on Haptics*, vol. 13, no. 2, pp. 368–379, 2020.

[167] E. Hoggan, S. A. Brewster, and J. Johnston, "Investigating the effectiveness of tactile feedback for mobile touchscreens," *Conference on Human Factors in Computing Systems - Proceedings*, pp. 1573–1582, 2008.

[168] Z. Ma, D. Edge, L. Findlater, and H. Z. Tan, "Haptic keyclick feedback improves typing speed and reduces typing errors on a flat keyboard," *IEEE World Haptics Conference, WHC 2015*, pp. 220–227, 2015.

[169] K. Tashiro, Y. Shiokawa, T. Aono, and T. Maeno, "Realization of button click feeling by use of ultrasonic vibration and force feedback," *Proceedings - 3rd Joint EuroHaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems, World Haptics 2009*, pp. 1–6, 2009.

[170] J. Monnoyer, E. Diaz, C. Bourdin, and M. Wiertlewski, "Ultrasonic friction modulation while pressing induces a tactile feedback," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9774, pp. 171–179, 2016.

[171] P. Garcia, F. Giraud, B. Lemaire-Semail, M. Rupin, and M. Amberg, "2motac: Simulation of button click by superposition of two ultrasonic plate waves," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12272 LNCS, pp. 343–352, 2020.

[172] P. Garcia, F. Giraud, B. Lemaire-Semail, M. Rupin, and A. Kaci, "Control of an ultrasonic haptic interface for button simulation," *Sensors and Actuators A: Physical*, vol. 342, no. January, 2022.

[173] D. Gueorguiev, A. Kaci, M. Amberg, F. Giraud, and B. Lemaire-Semail, "Travelling ultrasonic wave enhances keyclick sensation," in *Haptics: Science, Technology, and Applications* (D. Prattichizzo, H. Shinoda, H. Z. Tan, E. Ruffaldi, and A. Frisoli, eds.), (Cham), pp. 302–312, Springer International Publishing, 2018.

[174] M. Ercan Altinsoy, "Perceptual features of everyday push button sounds and audiotactile interaction," *Acoustical Science and Technology*, vol. 41, no. 1, pp. 173–181, 2020.

[175] S. Kaneko, T. Yokosaka, H. Kajimoto, and T. Kawabe, "A Pseudo-Haptic Method Using Auditory Feedback: The Role of Delay, Frequency, and Loudness of Auditory Feedback in Response to a User's Button Click in Causing a Sensation of Heaviness," *IEEE Access*, vol. 10, pp. 50008–50022, 2022.

[176] M. E. Altinsoy, "Perceptual aspects of auditory-tactile asynchrony," *Proceedings of the Tenth International Congress on Sound and Vibration*, pp. 3831–3838, 2003.

[177] Q. Hao, T. Ogata, K. I. Ogawa, J. Kwon, and Y. Miyake, "The simultaneous perception of auditory–tactile stimuli in voluntary movement," *Frontiers in Psychology*, vol. 6, no. 1429, pp. 1–10, 2015.

[178] T. Kaaresoja, S. Brewster, and V. Lantz, "Towards the temporally perfect virtual button: Touch-feedback simultaneity and perceived quality in mobile touchscreen press interactions," *ACM Trans. Appl. Percept.*, vol. 11, jun 2014.

[179] F. Karmali, S. E. Chaudhuri, Y. Yi, and D. M. Merfeld, "Determining thresholds using adaptive procedures and psychometric fits: evaluating efficiency using theory, simulations, and human experiments," *Experimental brain research*, vol. 234, no. 3, pp. 773–789, 2016.

[180] A. Vatakis, F. Balcı, M. Di Luca, and Á. Correa, *Timing and time perception: Procedures, measures, & applications*. Brill, 2018.

[181] R. Alcalá-Quintana and M. A. García-Pérez, "Fitting model-based psychometric functions to simultaneity and temporal-order judgment data: MATLAB and R routines," *Behavior Research Methods*, vol. 45, no. 4, pp. 972–998, 2013.

[182] J. Park, M. Schlag-Rey, and J. Schlag, "Voluntary action expands perceived duration of its sensory consequence," *Experimental brain research*, vol. 149, pp. 527–529, 2003.

[183] D. Brahimaj, F. Berthaut, F. Giraud, and B. Semail, "Cross-modal interaction of stereoscopy, surface deformation and tactile feedback on the perception of texture roughness in an active touch condition: Interaction intermodale de la stéréoscopie, de la déformation de surface et de la retour tactile sur la perception de la rugosité de la texture dans un état tactile actif," in *Proceedings of the 34th Conference on l'Interaction Humain-Machine*, pp. 1–12, 2023.

[184] R. A. Bolt, ""put-that-there": Voice and gesture at the graphics interface," *SIGGRAPH Comput. Graph.*, vol. 14, p. 262–270, July 1980.

[185] A. M. Fernandes and P. B. Albuquerque, "Tactual perception: A review of experimental variables and procedures," *Cognitive Processing*, vol. 13, no. 4, pp. 285–301, 2012.

[186] L. Vanacken, C. Raymaekers, and K. Coninx, "Evaluating the influence of multimodal feedback on egocentric selection metaphors in virtual environments," in *International Workshop on Haptic and Audio Interaction Design*, pp. 12–23, Springer, 2006.

[187] Y. Wang and C. L. MacKenzie, "Effects of orientation disparity between haptic and graphic displays of objects in virtual environments.," in *INTERACT*, vol. 99, pp. 391–398, 1999.

[188] G. Casiez, N. Roussel, R. Vanbelleghem, and F. Giraud, "Surfpad: riding towards targets on a squeeze film effect," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 2491–2500, 2011.

[189] N. A. Giudice, M. R. Betty, and J. M. Loomis, "Functional equivalence of spatial images from touch and vision: evidence from spatial updating in blind and sighted individuals.," *Journal of experimental psychology. Learning, memory, and cognition*, vol. 37 3, pp. 621–34, 2011.

[190] P. Olsson, F. Nysjö, S. Seipel, and I. Carlbom, "Physically co-located haptic interaction with 3d displays," in *2012 IEEE Haptics Symposium (HAPTICS)*, pp. 267–272, 2012.

[191] D. Picard, "Partial perceptual equivalence between vision and touch for texture information," *Acta Psychologica*, vol. 121, no. 3, pp. 227–248, 2006.

[192] W. M. Bergmann Tiest and A. M. Kappers, "Haptic and visual perception of roughness," *Acta Psychologica*, vol. 124, no. 2, pp. 177–189, 2007.

[193] G. Van Doorn, B. Richardson, D. Wuillemin, and M. Symmons, "Visual and haptic influence on perception of stimulus size," *Attention, perception and psychophysics*, vol. 72, pp. 813–22, 04 2010.

[194] Y. Ban, T. Narumi, T. Tanikawa, and M. Hirose, "Controlling perceived stiffness of pinched objects using visual feedback of hand deformation," in *2014 IEEE Haptics Symposium (HAPTICS)*, pp. 557–562, 2014.

[195] J. Kildal, "Tangible 3d haptics on touch surfaces: Virtual compliance," *Conference on Human Factors in Computing Systems - Proceedings*, pp. 1123–1128, 2011.

[196] R. L. Klatzky and S. J. Lederman, "Multisensory texture perception," in *Multisensory object perception in the primate brain*, pp. 211–230, Springer, 2010.

[197] S. J. Lederman and S. G. Abbott, "Texture perception: studies of intersensory organization using a discrepancy paradigm, and visual versus tactual psychophysics.," *Journal of Experimental Psychology: Human perception and performance*, vol. 7, no. 4, p. 902, 1981.

[198] Y.-X. Ho, M. S. Landy, and L. T. Maloney, "How direction of illumination affects visually perceived surface roughness," *Journal of vision*, vol. 6, no. 5, pp. 8–8, 2006.

[199] Z. Luo and A. Imamiya, "Do colors affect our recognition memory for haptic rough surfaces?," in *Computational Science - ICCS 2004* (M. Bubak, G. D. van Albada, P. M. A. Sloot, and J. Dongarra, eds.), (Berlin, Heidelberg), pp. 897–904, Springer Berlin Heidelberg, 2004.

[200] A. Lécuyer, "Simulating haptic feedback using vision: A survey of research and applications of pseudo-haptic feedback," *Presence: Teleoperators and Virtual Environments*, vol. 18, no. 1, pp. 39–53, 2009.

[201] Y. Ujitoko, Y. Ban, and K. Hirota, "Modulating Fine Roughness Perception of Vibrotactile Textured Surface using Pseudo-haptic Effect," *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 5, pp. 1981–1990, 2019.

[202] S. Günther, J. Rasch, D. Schön, F. Müller, M. Schmitz, J. Riemann, A. Matviienko, and M. Mühlhäuser, "Smooth as steel wool: Effects of visual stimuli on the haptic perception of roughness in virtual reality," in *CHI Conference on Human Factors in Computing Systems*, pp. 1–17, 2022.

[203] P. Punpongsanon, D. Iwai, and K. Sato, "¡italic¿softar¡/italic¿: Visually manipulating haptic softness perception in spatial augmented reality," *IEEE Transactions on Visualization and Computer Graphics*, vol. 21, no. 11, pp. 1279–1288, 2015.

[204] S. Han and J. Park, "Holo-Haptics: Haptic interaction with a see-through 3D display," *Digest of Technical Papers - IEEE International Conference on Consumer Electronics*, pp. 512–513, 2014.

[205] M. Dariosecq, P. Plénacoste, F. Berthaut, A. Kaci, and F. Giraud, "Investigating the semantic perceptual space of synthetic textures on an ultrasonic based haptic tablet," in *HUCAPP 2020*, (Valletta, Malta), Feb. 2020.

[206] R. Sahli, A. Prot, A. Wang, M. Müser, M. Piovarči, P. Didyk, and R. Bennewitz, "Tactile perception of randomly rough surfaces," *Scientific reports*, vol. 10, p. 15800, 09 2020.

[207] F. Argelaguet, D. A. G. Jauregui, M. Marchal, and A. LeCuyer, "Elastic images: Perceiving local elasticity of images through a novel pseudo-haptic deformation effect," *ACM Transactions on Applied Perception*, vol. 10, no. 3, pp. 1–14, 2013.

[208] T. Kawabe, "Mid-Air Action Contributes to Pseudo-Haptic Stiffness Effects," *IEEE Transactions on Haptics*, vol. 13, no. 1, pp. 18–24, 2020.

[209] T. Kimura and T. Nojima, "Pseudo-haptic feedback on softness induced by grasping motion," in *Haptics: Perception, Devices, Mobility, and Communication* (P. Isokoski and J. Springare, eds.), (Berlin, Heidelberg), pp. 202–205, Springer Berlin Heidelberg, 2012.

[210] S.-i. Yabe, H. Kishino, T. Kimura, and T. Nojima, "Pseudo-haptic feedback on softness induced by squeezing action," in *2017 IEEE World Haptics Conference (WHC)*, pp. 557–562, 2017.

[211] R. Hess, L. To, J. Zhou, G. Wang, and J. Cooperstock, "Stereo vision: The haves and have-nots," *i-Perception*, vol. 6, 2015.

[212] J. Wobbrock, L. Findlater, D. Gergle, and J. Higgins, "The aligned rank transform for nonparametric factorial analyses using only anova procedures," *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2011.

[213] C. T. Vi, D. Ablart, E. Gatti, C. Velasco, and M. Obrist, "Not just seeing, but also feeling art: Mid-air haptic experiences integrated in a multisensory art exhibition," *International Journal of Human Computer Studies*, vol. 108, no. June, pp. 1–14, 2017.

[214] M. O. Ernst and M. S. Banks, "Humans integrate visual and haptic information in a statistically optimal fashion," *Nature*, vol. 415, no. 6870, pp. 429–433, 2002.

[215] S. J. Lederman, "Tactile roughness of grooved surfaces: The touching process and effects of macro- and microsurface structure," *Perception & Psychophysics*, vol. 16, no. 2, pp. 385–395, 1974.

[216] T. C. Peck, L. E. Sockol, and S. M. Hancock, "Mind the Gap: The Underrepresentation of Female Participants and Authors in Virtual Reality Research," *IEEE Transactions on Visualization and Computer Graphics*, vol. 26, no. 5, pp. 1945–1954, 2020.

[217] R. A. Montano-Murillo, C. Nguyen, R. H. Kazi, S. Subramanian, S. DiVerdi, and D. Martinez-Plasencia, "Slicing-volume: Hybrid 3d/2d multi-target selection technique for dense virtual environments," in *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 53–62, IEEE, 2020.

[218] D. Brahimaj, E. Vezzoli, F. Giraud, and B. Semail, "Enhancing object localization in vr: Tactile-based hrtf and vibration headphones for spatial haptic feedback," *IEEE Transactions on Haptics (ToH)*, 2024.

[219] C. C. Berger, M. Gonzalez-Franco, A. Tajadura-Jiménez, D. Florencio, and Z. Zhang, "Generic hrtfs may be good enough in virtual reality. improving source localization through cross-modal plasticity," *Frontiers in neuroscience*, vol. 12, p. 21, 2018.

[220] A. Ahrens, K. D. Lund, M. Marschall, and T. Dau, "Sound source localization with varying amount of visual information in virtual reality," *PloS one*, vol. 14, no. 3, p. e0214603, 2019.

[221] J. B. V. Erp, H. A. V. Veen, C. Jansen, and T. Dobbins, "Waypoint navigation with a vibrotactile waist belt," *ACM Transactions on Applied Perception (TAP)*, vol. 2, no. 2, pp. 106–117, 2005.

[222] W. Heuten, N. Henze, S. Boll, and M. Pielot, "Tactile wayfinder: a non-visual support system for wayfinding," in *Proceedings of the 5th Nordic conference on Human-computer interaction: building bridges*, pp. 172–181, 2008.

[223] I. Peate, "The skin: largest organ of the body," *British Journal of Healthcare Assistants*, vol. 15, no. 9, pp. 446–451, 2021.

[224] S. Paneels, M. Anastassova, S. Strachan, S. P. Van, S. Sivacoumarane, and C. Bolzmacher, "What's around me Multi-actuator haptic feedback on the wrist," *2013 World Haptics Conference, WHC 2013*, pp. 407–412, 2013.

[225] A. Meier, D. J. Matthies, B. Urban, and R. Wettach, "Exploring vibrotactile feedback on the body and foot for the purpose of pedestrian navigation," *ACM International Conference Proceeding Series*, vol. 25-26-June-2015, 2015.

[226] N. K. Dim and X. Ren, "Investigation of suitable body parts for wearable vibration feedback in walking navigation," *International Journal of Human Computer Studies*, vol. 97, pp. 34–44, 2017.

[227] O. B. Kaul and M. Rohs, "HapticHead: A spherical vibrotactile grid around the head for 3D guidance in virtual and augmented reality," *Conference on Human Factors in Computing Systems - Proceedings*, vol. 2017-January, pp. 3729–3740, 2017.

[228] H. Kerdegari, Y. Kim, and T. J. Prescott, "Head-mounted sensory augmentation device: Comparing haptic and audio modality," in *Biomimetic and Biohybrid Systems* (N. F. Lep-

ora, A. Mura, M. Mangan, P. F. Verschure, M. Desmulliez, and T. J. Prescott, eds.), (Cham), pp. 107–118, Springer International Publishing, 2016.

[229] R. L. Peiris, W. Peng, Z. Chen, and K. Minamizawa, "Exploration of cuing methods for localization of spatial cues using thermal haptic feedback on the forehead," *2017 IEEE World Haptics Conference, WHC 2017*, no. June, pp. 400–405, 2017.

[230] R. L. Peiris, W. Peng, Z. Chen, L. Chan, and K. Minamizawa, "ThermoVR: Exploring integrated thermal haptic feedback with head mounted displays," *Conference on Human Factors in Computing Systems - Proceedings*, vol. 2017-May, pp. 5452–5456, 2017.

[231] M. Rietzler, K. Plaumann, T. Kräenzle, M. Erath, A. Stahl, and E. Rukzio, "VaiR: Simulating 3D airflows in virtual reality," *Conference on Human Factors in Computing Systems - Proceedings*, vol. 2017-May, pp. 5669–5677, 2017.

[232] Y. Silina and H. Haddadi, "New directions in jewelry: A close look at emerging trends & developments in jewelry-like wearable devices," *ISWC 2015 - Proceedings of the 2015 ACM International Symposium on Wearable Computers*, no. September, pp. 49–56, 2015.

[233] H. Gil, H. Son, J. R. Kim, and I. Oakley, "Whiskers: Exploring the use of ultrasonic haptic cues on the face," *Conference on Human Factors in Computing Systems - Proceedings*, vol. 2018-April, pp. 1–13, 2018.

[234] M. Lee, S. Je, W. Lee, D. Ashbrook, and A. Bianchi, "ActivEarring: Spatiotemporal Haptic Cues on the Ears," *IEEE Transactions on Haptics*, vol. 12, no. 4, pp. 554–562, 2019.

[235] UnityTechnologies, "Unity, https://unity3d.com/," 2023.

[236] Meta, "Oculus quest 2, https://www.oculus.com/quest-2," 2020.

[237] Unity, "Audio spatializer sdk," 2023.

[238] N. W. Hochreiter, M. J. Jewell, L. Barber, and P. Browne, "Effect of vibration on tactile sensitivity," *Physical Therapy*, vol. 63, no. 6, pp. 934–937, 1983.

[239] M. Halldestam, "Anova-the effect of outliers," 2016.

Laboratoire d'électrotechnique et d'électronique de puissance de Lille

Titre: Intégration de la Rétroaction Haptique dans les Appareils Intelligents : Interfaces Multimodales et Directives de Conception

Mots clés: Haptic Feedback - Synchronisation sensorielle - Interface multimodale - Perception

Résumé: L'intérêt croissant pour l'intégration de la rétroaction haptique dans les produits commerciaux est directement lié aux progrès de la technologie haptique. Notamment, la prolifération des smartphones et des tablettes a conduit à l'intégration de modalités haptiques pour diverses fonctions. Alors que des recherches approfondies ont exploré l'intégration des modalités sensorielles (visuelle, auditive, tactile) dans le toucher passif, il existe un manque relatif de connaissances en ce qui concerne la bimodalité ou la multimodalité dans le contexte du toucher actif. Les technologies émergentes, telles que l'haptique de surface, offrent des opportunités pour étudier divers aspects liés à l'intégration sensorielle.

Ce travail fournit des lignes directrices précieuses pour les développeurs, tirées d'études expérimentales dans le domaine du toucher actif. Notre première investigation se concentre sur la relation temporelle entre les retours audio et tactiles, révélant un seuil critique de 200 ms lors des interactions de glissement sur une surface haptique. De plus, nous identifions un délai audio-tactile acceptable de 109 ms pour les gestes de clic avec des boutons virtuels, soulignant la nécessité de prohiber ou de minimiser le délai haptique à moins de 40 ms. Une étude comparative impliquant des individus voyants et aveugles dévoile un aspect crucial de l'inclusion : le respect des limites de synchronisation audio-tactile de la population voyante, concerne les boutons virtuels, permet la conception inclusive d'interfaces adaptées aux deux populations. De plus, nous explorons l'impact de facteurs tels que la stéréoscopie et la déformation de surface sur la perception de la rugosité des textures, démontrant que leur présence peut altérer la rugosité perçue des textures lisses de plus de 20%.

En outre, notre recherche explore le potentiel de l'utilisation de casques vibrants pour la localisation d'objets, révélant une sensibilité de 7° pour la modalité haptique, de 8° pour la rétroaction auditive et de 6° pour la rétroaction audio-tactile. Cela met en évidence non seulement la viabilité de la rétroaction haptique en réalité virtuelle pour la localisation d'objets, mais aussi l'amélioration obtenue en renforçant l'expérience sensorielle avec des stimuli audio-tactiles.

Université de Lille
L2EP / IRCICA
50 Avenue Halley, 59650 Villeneuve-d'Ascq, France

Université de Lille

Laboratoire d'électrotechnique et d'électronique de puissance de Lille

Title: Integrating Haptic Feedback in Smart Devices: Multimodal Interfaces and Design Guidelines

Keywords: Haptic Feedback - Sensory synchronization - Multimodal interface - Perception

Abstract: The growing interest in integrating haptic feedback into commercial products is a direct result of advancements in haptic technology. Notably, the proliferation of smartphones and tablets has led to the integration of haptic modalities for various interfaces.
While extensive research has explored the integration of sensory modalities (visual, auditory, tactile) in passive touch, there is a relative dearth of knowledge regarding bimodality or multimodality in the context of active touch. Emerging technologies, like surface haptics, offer opportunities to investigate various aspects related to sensory integration. This work provides valuable guidelines for developers, drawing from experimental studies in the realm of active touch. Our initial investigation focuses on the temporal relationship between audio and tactile feedback, revealing a critical 200 ms threshold during sliding interactions on a haptic surface. Moreover, we identify an acceptable audio-tactile delay of 109 ms for click gestures with virtual buttons, emphasizing the need to prohibit or minimize haptic delay to less than 40 ms. A comparative study involving sighted and blind individuals unveils a crucial aspect of inclusion: adhering to synchronization boundaries of the sighted population, relative to virtual buttons, allows for the inclusive design of interfaces accommodating both populations. Additionally, we delve into the impact of factors such as stereoscopy and surface deformation on the perception of texture roughness, demonstrating that their presence can alter the perceived roughness of smooth textures by over 20%. Furthermore, our research explores the potential of using vibration headphones for object localization, revealing a sensitivity of 7° for the haptic modality, 8° for auditory feedback, and 6° for audio-tactile. This highlights not only the viability of haptic feedback in virtual reality for object localization but also the improvement achieved by reinforcing the sensory experience with audio-tactile stimuli.

Université de Lille
L2EP / IRCICA
50 Avenue Halley, 59650 Villeneuve-d'Ascq, France

Université de Lille