Continuous auditory feedback for sensorimotor learning

Eric Boyer

To cite this version:

Eric Boyer. Continuous auditory feedback for sensorimotor learning. Cognitive Sciences. Université Pierre et Marie Curie - Paris VI, 2015. English. NNT : 2015PA066165 . tel-01191826v2

HAL Id: tel-01191826

https://hal.sorbonne-universite.fr/tel-01191826v2

Submitted on 13 Nov 2015

Thèse de Doctorat de l'Université Pierre et Marie Curie
École Doctorale ED3C

Spécialité
Sciences cognitives

Présentée et soutenue par

Éric Boyer

Pour obtenir le grade de
Docteur de l'Université Pierre et Marie Curie

Continuous Auditory Feedback for Sensorimotor Learning

Thèse dirigée par
Patrick Susini et Sylvain Hanneton

et encadrée par
Frédéric Bevilacqua

préparée à l'Ircam - STMS CNRS UPMC
et au LPP UMR8242, Université Paris Descartes

soutenue le 11 mai 2015 devant un jury composé de :

Patrick SUSINI	Ircam	Directeur
Sylvain HANNETON	Université Paris Descartes	Directeur
Frédéric BEVILACQUA	Ircam	Encadrant
Cathy CRAIG	Université Queen's de Belfast	Rapporteur
Roberto BRESIN	KTH Stockholm	Rapporteur
Vincent HAYWARD	UPMC	Examinateur
Bruno GIORDANO	Université de Glasgow	Examinateur
Olivier GAPENNE	Université de Compiègne	Examinateur

Abstract

Our sensorimotor system has developed a strong relationship with the auditory space surrounding us. We have implicitly learned to integrate the sonic outcomes of our actions and use them everyday as auditory feedback. The development of motion sensing and audio technologies allows for designing auditory feedback, by creating specific interaction between sound and motion. In this PhD research, we focus on continuous movement *sonification*, the generation of synchronous sound feedback controlled by movement or action features.

Auditory feedback in the action-perception loop has received considerably less attention compared to visual feedback. Moreover, the learning mechanisms occurring in interactive systems using sound feedback, such as digital musical interfaces, have been rarely studied. In this thesis, we evaluate the contribution of continuous auditory feedback to sensorimotor control and learning in interactive systems. To fulfill this aim, we consider various sonic interaction setups (visual/non-visual, tangible/virtual), that we use in five experimental studies, targeting the questions of *how* the feedback can be integrated and *which* gesture-sound parameters are relevant to learning.

First, considering a pointing task, we observe that the auditory system integrates dynamic auditory cues for online motor control. Auditory representations of space and of the scene can be built from audio features and transformed into motor commands. Second, during the exploration of a virtual object solely through sound, we observe that the emerging auditory-motor representations can also shape exploratory movements and allow for tactile to auditory sensory substitution.

Third, we measure that continuous auditory feedback in a tracking task helps significantly the performance. Both error and task sonification improve tracking but have different effects on learning. The sonification of the user's movement can increase the energy of the produced motion and prevent feedback dependency. We also tested a system enabling the sonification of eye tracking movements. The auditory feedback had important effects on eye movements control, both positively and negatively across the participants, during both guided and free pursuit.

Finally, we present the concept of sound-oriented task, where the target is expressed as acoustic features to match. We show that users can achieve this auditory task with various responsiveness to the system, and that motor adaptation can be driven by interactive audio cues only. This context illustrates the potential of placing the sound at the core of the interaction task for integration in the sensorimotor loop.

Globally, our results show the benefits of continuous movement sonification for motor learning, specifically on gestural features, performance and energy. We can derive design principles from our studies, such as the use of 'ecological' gesture-sound mapping. In particular, we propose to consider a 'moving while listening' approach where the user learns to associate sound features with corresponding movement features. Among the potential applications, we believe this approach can be successfully applied in physical rehabilitation and neurodegenerative diseases.

Résumé

Notre système sensorimoteur a développé une relation particulière avec son environnement sonore et utilise constamment les conséquences acoustiques de nos actions, que nous percevons comme un retour sonore. Le développement des systèmes de captation gestuelle et des technologies audio permet de manipuler ce retour sonore, notamment en créant une interaction entre le son et le geste, dont le couplage peut alors s'effectuer de nombreuses manières. Dans ce travail de doctorat, nous proposons d'étudier la sonification gestuelle continue, c'est-à-dire la génération d'un retour sonore synchrone durant le geste, dont les paramètres acoustiques dépendent du mouvement produit ou de l'action.

La grande majorité des études sur la boucle perception-action s'est focalisée sur le flux visuel et peu sur l'utilisation d'un retour sonore. En outre, les mécanismes d'apprentissage mis en œuvre dans les systèmes interactifs sonores sont rarement étudiés. Le travail que nous présentons a pour but d'évaluer la contribution d'un retour sonore continu à l'apprentissage et au contrôle moteur dans des systèmes interactifs. Dans cet objectif, nous considérons plusieurs situations interactives sonores (avec ou sans vision, tangible/virtuel), que nous utilisons lors de cinq études expérimentales. Les questions concernent l'identification des mécanismes d'intégration du retour sonore et des paramètres geste-son pertinents pour l'apprentissage.

Premièrement, nous observons lors d'une tâche de pointage que le système auditif intègre en continu des indices acoustiques dynamiques pour le contrôle moteur. Des représentations sonores de l'espace peuvent être construites à partir de ces informations et transformées en commandes motrices permettant de corriger le mouvement. Deuxièmement, dans le cas de l'exploration d'un objet virtuel sonore, nous observons que les représentations audio-motrices émergentes peuvent également influencer des mouvements exploratoires et permettre des cas de substitution sensorielle, notamment un transfert auditif vers des sensations tactiles et proprioceptives.

Troisièmement, nous montrons dans le cas d'une tâche de suivi visuo-manuel qu'un retour sonore continu peut améliorer significativement la performance. Les sonifications apportant un retour relatif à l'erreur de suivi (distance) ainsi qu'à la tâche (comportement de la cible) permettent d'améliorer la qualité du suivi, mais montrent des effets distincts sur l'apprentissage. Nous montrons également que la sonification du mouvement de l'utilisateur peut augmenter l'énergie du geste produit et contribue à un apprentissage robuste, non dépendant du retour sonore additionnel. Nous proposons également un système permettant la sonification en temps-réel des mouvements de poursuite oculaire. La sonification de tels mouvements montre d'importants effets, sur le contrôle oculomoteur (qu'il soit libre ou en suivi guidé), à la fois positifs et négatifs selon les participants.

Enfin, nous présentons le concept de “tâche sonore”, dans lequel la cible est présentée et s'exprime sous forme de paramètres sonores à reproduire. L'expérience montre que les utilisateurs apprennent à effectuer cette tâche, avec toutefois des disparités dans leurs aptitudes. Nous observons

également qu'une adaptation motrice peut être induite par des changements dans le mapping geste-son, modifiant l'objectif sonore à atteindre. Ce contexte met en évidence le potentiel de l'information sonore replacée au cœur de l'interaction pour l'intégration dans la boucle sensorimotrice.

Dans leur ensemble, nos résultats montrent les effets bénéfiques de la sonification du mouvement sur l'apprentissage moteur, notamment sur la performance et la dynamique du geste. Il est possible de tirer de nos études des principes de design de l'interaction, comme par exemple l'utilisation de mappings geste-son 'écologiques'. Nous proposons particulièrement de considérer une approche de "l'écoute du mouvement", dans laquelle l'utilisateur apprend à associer des indices audio avec les caractéristiques intrinsèques de ses mouvements. Ce type d'interaction permet d'envisager diverses applications dans des cas d'apprentissage moteur. Il semble en effet que cette approche puisse être appliquée avec succès à la rééducation physique de l'hémiparésie ou de maladies neurodégénératives.

Contents

À mes parents, à mon frère

Chapter 1

Introduction

1.1 General introduction

Sound is everywhere in our environment, sometimes desired, sometimes unwanted, but always informs about an event. Every action or moving object leads to the creation of a related sound wave, as an acoustical and dynamical footprint of the action. The resulting sound carries spatial cues and information about movement that we can perceive, whether the action originates from ourself or from the surrounding environment. Throughout our development, we have built a particular sensibility to sounds that are produced by our actions or others'. Our brain learns unconsciously to relate our actions with the resulting auditory feedback, and so much that some task become difficult if we cannot *hear* them, like driving a car or playing table tennis. Yet, little is known about the integration mechanisms of auditory feedback in during interactions. The psychology and motor control fields rarely explore action-sound relationships in this context. Besides, the digital musical instrument community, which developed advanced gestural audio control techniques, relies on idiosyncratic design and does not take into account explicitly sensorimotor learning mechanisms. Whereas much research emphasizes the potential benefits of multisensory learning and auditory feedback, interactive sonification has been receiving the deserved attention only recently (Dubus, 2013).

The general aims of this work include investigating the role of auditory feedback in gestural interactive systems through experimental studies, in order to address fundamental questions. We particularly focus on how continuous auditory feedback can be beneficial for movement control and learning, and explore gesture-sound relationships in interactive scenarios. We also consider important applications, specifically physical rehabilitation whose systems and procedures could benefit from sound interaction from the point of view of motor learning. Interactive sound design, for sonified object manipulation, musical interfaces and sport pedagogy are also targeted. Our work intends to contribute to the emergence of fundamental principles of movement sonification for learning, which are scattered over various disciplines today. We propose use-case scenarios and concepts exploring the idea of "moving while listening": continuous sonic interactions to study auditory-motor learning.

We believe indeed that a better understanding of the auditory-motor loop in gestural interactive systems can bring, on the one hand, valuable experimental contexts and paradigms to address fundamental questions, and on the other hand, platforms for the development of important applications such as physical rehabilitation. This document first presents important results from the background disciplines, relative to the use of auditory feedback. Then, five experimental studies are reported focusing on continuous movement sonification, each one applied to a particular sensorimotor task. Two of them, visuo-manual tracking and pointing, are standard sensorimotor tasks that we study in a sonic context. The other three present new frameworks of movement sonification in unusual contexts: eye pursuit control, gestural sonic interaction with a virtual object, and motor adaptation to a sound target in an audio only task.

1.2 Research context

The present work has been prepared during three years of research, in the framework of the LEGOS project[1]. This project gathered research teams from Ircam (Science Technology of Music and Sound lab, CNRS UPMC) and from the Psychology of Perception lab (LPP - CNRS UMR 8242) from Paris Descartes University. It aimed at crossing expertise on gesture-sound control technologies for addressing issues about sensorimotor learning. Specifically, although available gestural interfaces and audio technology are quite advanced, the design of interactive systems rarely takes sensorimotor and learning aspects into account. The objectives of the project were to open lines of discussion on how connected fields could inspire but also benefit from these systems, and to develop interaction prototypes for sensorimotor learning in various contexts. Expertise on gestural control of media and audio technology was brought by Ircam, whereas LPP brought sensorimotor and rehabilitation insights. This work has been carried out at both Ircam (with the help of the Sound, Music and Motion Interaction team, as well as Sound Design and Perception team) and LPP.

Our main experimental apparatus (and background) is gesture-sound interactive systems, integrating motion sensing and real-time audio technologies from Ircam. In the following paragraphs, we will describe the architecture of the systems we propose to use and cite supportive works from neurophysics and cognition sciences on the action-sound relationship.

1.2.1 Gesture-sound interactive systems

The development of gestural interfaces in human-machine interaction, video games and mobile platforms brought new devices and usages, where gestures play a central role. These new usages developed extremely fast and now cover a wide and still growing range of

---

[1] ANR-11-BS02-012

technological applications, engaging the user in a gestural relationship with the system. Typical examples are the gesture-based and whole body video games, tactile mobile platforms, domestic electronics or musical interfaces. Digital media, like sounds, can now be produced and controlled in real-time by gestural input.

A gesture-sound interactive system can be basically described by the flow chart represented on figure 1.1. We depict such a system as engaging a "computer mediated interaction" between the user and sound. As the user performs an action, either with an interface or not, his gestures are captured and analyzed. The extracted movement features are then utilized to control a sound synthesis engine. This step, the 'mapping' step, plays a central role in the system, as it defines how the acoustic features of the produced sound will be related to motion. The mapping can be a very simple numerical function or a complex dynamic system like a machine learning unit. It can link one gestural feature to one sound parameter, or one to many, many to one or many to many[2]. The output sound is perceived by the user, primarily as an auditory outcome (feedback) of his action. The resulting action/perception loop, in addition to the intrinsic feedback the user receives (*e.g.* proprioceptive or vestibular), allows for creating an "embodied" interaction between movement and sound.

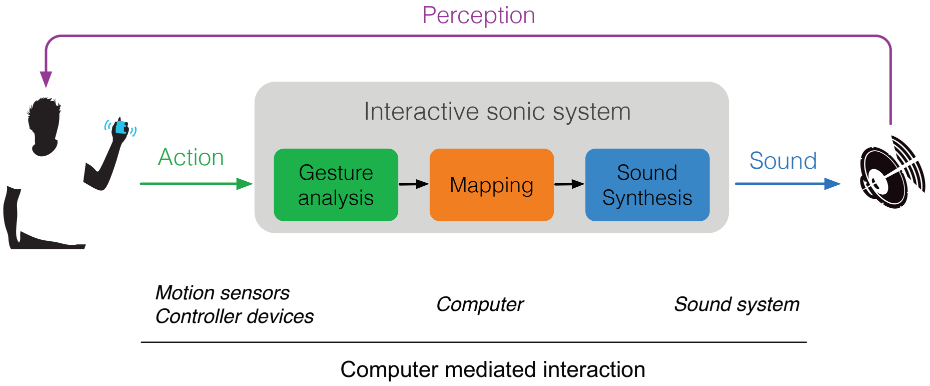


Figure 1.1 – Schematic of a gesture-sound interactive system.

Obviously, time flow is a key element of the interaction that needs to be accurately preserved. Both high-end and low-cost motion capture systems (like Microsoft Kinect, Leap Motion or embedded sensors in mobile devices) now offer adequate temporal performances. Real-time digital audio treatment and synthesis (*e.g.* Max/MSP platform or SuperVP technology[3]) allow for seamless control and extensive design possibilities.

1.2.2 Why auditory feedback?

Our brain deeply integrates the causal relationship between an action and the resulting sound. Research has brought results which show both neural and behavioral aspects of this

---

[2] The vocabulary comes from the specific field.

[3] `http://anasynth.ircam.fr/home/english/software/supervp`

link, and can lay the foundation for the benefits to use auditory feedback for movement learning.

Physiologically speaking, the ear is quite powerful to discriminate fine temporal events, like amplitude and frequency modulations, particularly compared to vision (Robertson et al., 2009). Processing time of auditory information can be shorter than visual (Baram and Miller, 2007). Besides, auditor information can be processed by our sensory system without interfering with normal visual or proprioceptive feedback processes (Robertson et al., 2009), allowing for creative multi-sensory interactions. Vision allows our system to make previsions of approaching events and to anticipate in a dynamical environment, but the benefits of multimodal, auditory-visual, interactions are generally emphasized (Shams and Seitz, 2008). Auditory-visual interactions can help perception (Bulkin and Groh, 2006), and learning new skills (Seitz et al., 2006). In addition, the auditory modality can be integrated by sensory-impaired people (Dozza et al., 2007). Audition is also an efficient channel for providing spatial information, as it is constantly used by our positioning systems (Maier and Groh, 2009) and to calibrate body perception (Tajadura-Jiménez et al., 2012).

In the mid 90's, a particular class of visuomotor neurones has been discovered in the pre-motor cortex of monkeys: the mirror neurons, which fire when an action is produced, but also when the monkey observes someone performing the action with on object. These neurones do not fire without the object present (mime) or without the action being performed (Rizzolatti et al., 2001; Rizzolatti and Craighero, 2004). The mirror system, also detected in humans, is thought to use auditory information as well. Auditory-visual mirror neurons have been detected and seem to code the meaning of actions whether the action is performed, heard or seen (Kohler et al., 2002). This mechanism allows for internal representation of an auditory object when hearing the sound of an action, linked to the source and to the context (De Lucia et al., 2009). The motor system has also been showed to 'embody' higher level representations of perceived actions, encoded in auditory, visual and motor domains (Ticini et al., 2012). These representations allow for increased motor excitability after hearing the sound of previously learned voluntary actions, and appear to match the final goal of the action. Hearing action sounds has been shown to influence self perception as well (Tajadura-Jiménez et al., 2012), indicating that body representations integrate auditory cues.

These specific works confirm, at a deep neural level of course, the close relationship between auditory input and active actions. They provide neural basis encouraging to investigate auditory feedback for action-related tasks and behavioral assessments.

1.3 Specific aims of the work

The present work focuses on particular sensorimotor situations involving motor tasks which are usually accomplished utilizing visual feedback. Here, we study them with either replacing or augmenting feedback inputs with auditory feedback, through gesture-sound interactive systems we developed. Depending on the context and the tasks, the auditory feedback is thought to improve sensorimotor performance or to be an essential learning parameter, participating to a multi-sensory experience. Various papers supporting this idea and providing promising results will be described in the present work.

Our research questions are structured according to the role played by the auditory feedback in the task. On the one hand, sound is considered as a medium to bring supplementary information, in the case of task which need to be performed with visual input. On the other hand, we propose situations where sound plays a central role, without which the action cannot be performed. These scenarios will allow us to explore new usage of auditory feedback and movement sonification, in order to address sensorimotor learning in interactive contexts. Specifically, we will address how continuous auditory feedback is integrated in the sensorimotor loop and measure its effects on performance and learning. Over the five studies we present in this document, we formalized the following questions:

- How is auditory information integrated into the sensorimotor loop during a spatial pointing task?
- Can auditory interaction enable sensory substitution, leading to sense a virtual object in space?
- Can continuous auditory feedback improve performance and learning in a continuous motor task, either a well-known action or a new one?
- Is it possible to observe motor learning and adaptation in a sonic context, where sound represents both the interaction and the final goal?

Through our different studies, we will underline both experimental opportunities given by gesture-sound interaction, but also consider important applications. Auditory-motor behavior will be assessed through performance evaluation and movement (or auditory output) features, both provided within the system architecture (sensors, interfaces, etc.).

1.4 Experimental choices

Based on the wide range of related scientific literature, there is a need to focus on the common interests of utilizing sound in different disciplines (psychology, motor control, arts, rehabilitation, sports, etc.). Our work, which intends to contribute to both fundamental knowledge and applications, explores different contexts and tasks. Several experimental guidelines were established to enlarge the potential fields of applications.

1.4.1 Materials and methods

The evaluation of learning and performance was carried out using quantitative movement analysis and behavioral assessments (mainly questionnaires and observations). We employed mostly custom inertial measurement units (accelerometers, gyroscopes), or commercial motion tracking devices, so that the systems can be of reasonable size and easily reproduced[4]. Using 'low-cost' sensors and motion capture devices also allows to consider mobile and domestic applications. The available sensing technology and widespread, fair quality, home sound systems, favor reasonable cost and home-based applications.

1.4.2 Sound synthesis

The relative simplicity of auditory feedback described in the literature encourages to take inspiration from music technologies (Bevilacqua et al., 2013b), which offer remarkable sound control and synthesis techniques. Nonetheless, as we address sensorimotor learning in this research work, the feedback message need to be easily accessible so that subjects can build auditory-motor representations, with a minimum conscious cognitive load. As a result we designed sound interactions with direct perceivable characteristics, while offering rich temporal and spectral audio features. Similarly, we chose to use sound materials avoiding aesthetic connotation as much as possible. One of the aim was to limit emotional response that could influence behavior by the subjective meaning sound can carry. Assuming it is possible to limit the emotional response of a user in an interactive system, we will discuss this choice in the final chapter, section 8.2.

Consequently, we developed our sonification processes avoiding musical samples and natural sounds; the particular case of the effects of music and natural sounds on mind and body is indeed a whole branch of neuroscience.

We also chose to avoid rhythmic and sequential feedback (also subject of specific studies) to emphasize with sound the continuous aspect of movement. More generally, we target applications in various domains of interactions, using different interfaces. To this aim, we focused on embodied, 'physical' sonic interactions, and kept away from the determinism of musical sounds.

Designing direct interaction, demanding no particular cognitive capacities, would also make the systems accessible to cognitively impaired populations. We particularly think of mental impaired, autistic users, but also children, or neurodegenerative disorder patients (Alzheimer or Parkinson's diseases).

From a practical design point of view, the sound models for movement sonification were chosen taken the previous elements into account and in view of illustrating as simply as possible time variations of gesture parameters and outcomes. With this aim, the acoustical features of the resulting sound should allow to identify clearly changes, with a fine precision and within a wide dynamics range. Consequently, we considered mostly wide-band noises which we processed through various filters, driven by movement or task parameters. An additional layer of audio treatments was often used to ensure minimum aesthetics and pleasantness, mainly preventing audio artifacts (buzz, low-frequency noise) and adapting the sound to the task considered. This technique allows for accessible, clear and responsive

---

[4] The use of cerebral imaging or physiological measurements, like EEG or EMG, was not addressed in this work

audio illustration of movement features, while operating with low CPU resource.

As far as physical models are concerned, they offer the possibility to generate natural sounds from materials (such as wood, metal or glass), with a high quality sound rendering and a large synthesis parameters choice. However, the interaction possibilities and the sometimes complex controls make them difficult to integrate in experimental setups. In addition, choosing a specific material might give the user some clues about the object he is interacting with, or referring to some specific physical properties or affordances, introducing bias in the protocols. In the case of virtual object manipulation (chapter 4) for instance, the use of physical model might proved disturbing as no tangible object is manipulated.

1.4.3 Proposed concepts

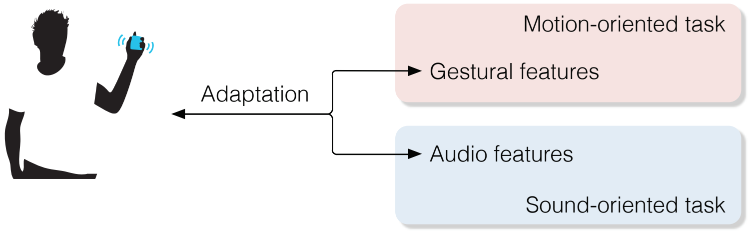


Figure 1.2 – Motion-oriented and sound-oriented task concepts.

We propose to consider the action-perception coupling in sonic interactive systems by conceptualizing two types of task, blooming the specificities of interacting with sound. The loop in which the user is engaged can be seen from two different angles. If the aim of the user is to accomplish a particular motion, the interactive system generates sound as a *feedback* on the resulting movements and actions. But the system can also provide the user with the capacity to trigger and/or control a sound, or an acoustic feature of a sound, from a given gesture. In that case, the goal is to produce a specific sound, which will result from the input action. Depending on how the goal is expressed, in terms of movement or acoustic features, we define the task as either “motion-oriented” or “sound-oriented” (figure 1.2). The limit between the two seems narrow, also because the approaches share the same system architecture. But the importance lies in the specific *goal* for the user, and changes the way he understands the task and controls his attention. Specifically we emphasize that a task can be defined as sound-oriented, focus the attention on audio features and without any musical context. Different designs and protocols can result from this view (an example[5] will be given in chapter 7). The classic tasks we study here in a sonic interactive context will also bring results supporting this view of the interaction between sound and motion. Although the large majority of the literature refers to motion-oriented tasks, we propose to consider sonic interaction, particularly movement sonification, as a whole framework for sensorimotor learning experiments.

---

[5] Other experiments from this work could suit this view too.

1.5 Structure of the manuscript

We first introduce in chapter 2 basic notions and main theoretical aspects of sensorimotor learning, based on the available literature. In a second part of this chapter, we report on experimental results regarding the use of auditory feedback and sonification for motor learning. Finally, we approach the main connected fields, which bring either background results or important application field (Sonic Interaction Design, music, sports and rehabilitation). Although we tried to be as comprehensive as possible in citing previous important works, we do not claim we present here an exhaustive review on the subject; firstly because the multiple fields involved scatter the results over different approaches and disciplines, secondly to ensure this document is not too long and thus remains pleasant to read.

The five following chapters present five experimental works, either published, submitted or exploratory work, meant to be completed and published in a near future. In these studies, continuous movement sonification represents either a feedback on movements produced, or a key element of the learning process.

Chapter 3 addresses the role of auditory feedback in a pointing task to virtual auditory sources. Experimental results provide evidence for an online motor regulation enabled by auditory representations and illustrates fundamental aspects of the auditory-motor loop. Representations are also addressed in chapter 4, where we develop a concept of gestural sonic interaction with a virtual object. This chapter also contains a sensory substitution study, exploring the relationships between touch, proprioception and sound. Chapters 5 and 6 focus on tracking tasks, which have been quite studied in motor control, but rarely with auditory feedback. Chapter 5 presents an experimental study of tracking with continuous auditory feedback, comparing three types of sonification. Chapter 6 presents an innovative experimental platform and protocol of sonification for eye movements control. It highlights a new application of movement sonification, whose exploration is only beginning. In chapter 7, we develop the concept of sound-oriented task where the goal and learning parameters consist in audio feature (as opposed to the previous chapter describing movement-oriented tasks). This study addresses the issue of learning a movement with an audio target.

While chapters 3 and 4 illustrate fundamental aspects of the auditory-motor loop, chapters 5 and 6 show how continuous movement sonification participates to sensorimotor learning in a tracking exercise, and chapter 7 shows that considering gestural sonic interaction as a whole experimental framework is possible. Finally, we will discuss and summarize the results of the present work in chapter 8, where we highlight the fundamental results and important applications this work provides.

1.6 Contributions

Peer-reviewed journal articles and book chapters

E. O. Boyer, F. Bevilacqua, S. Hanneton. “Continuous sound feedback for motor learning: investigating the auditory-motor loop with interactive sonification on a visuo-manual tracking task”, Submitted to *Learning and Memory.*

E. O. Boyer, Q. Pyanet, S. Hanneton, F. Bevilacqua. “Learning Movement Kinematics with a Targeted Sound”. *In Sound, Music & Motion, Volume 8905 of Lecture Notes in Computer Science*, Springer. 2014

E. O. Boyer, B. M. Babayan, F. Bevilacqua, M. Noisternig, O. Warusfel, A. Roby-Brami, S. Hanneton, I. Viaud-Delmon. “From ear to hand: the role of the auditory-motor loop in pointing to an auditory source”. *Frontiers in Computational Neuroscience*, 7(April), 26. 2013

E. O. Boyer, F. Bevilacqua, F. Phal, S. Hanneton. “Low-cost motion sensing of table tennis players for real time feedback”. *International Journal of Table Tennis Sciences*, 8. 2013

Peer-reviewed international conferences

A. Roby-Brami, A. Van Zandt-Escobar, N. Jarrassé, J. Robertson, N. Schnell, E. O. Boyer, S. Hanneton, Bevilacqua, F. “Toward the use of augmented auditory feedback for the rehabilitation of arm movements in stroke patients”. *Annals of Physical and Rehabilitation Medicine*, 57, e4–e5. 2014

F. Bevilacqua, A. Van Zandt-Escobar, N Schnell, E. O. Boyer, N. Rasamimanana, S. Hanneton, A. Roby-Brami. “Sonification of the Coordination of Arm Movements”. *In Proceedings of Multimodal Behavior: Impact of Sound.* Hannover, Germany. 2013

E. O. Boyer, L. Colin Malagon, F. Bevilacqua, P. Susini, S. Hanneton. “Continuous Sound Feedback in Tracking Tasks”. *In Proceedings of Multimodal Behavior: Impact of Sound.* Hannover, Germany. 2013

F. Bevilacqua, E. Fléty, N. Schnell, B. Caramiaux, N. Rasamimanana, J. Françoise, J. Bloit, E. O. Boyer. “De-MO: Designing Action-Sound Relationships with the MO Interfaces”. *In CHI’13 Extended Abstracts on Human Factors in Computing Systems*, pp. 2907-2910. 2013

E. O. Boyer, Q. Pyanet, S. Hanneton, F. Bevilacqua. “Guiding Motion using a Sound Target”. *In Proceedings of the 10th International Symposium on Computer Music Multidisciplinary Research, Sound, Music & Motion, CMMR2013*, pp. 176–89. Marseille, France. 2013

Other scientific communications

E. O. Boyer. “Touching Sounds”. *Audiomostly 2014*, PhD symposium, Allborg, Denmark. 2014

E. O. Boyer, Q. Pyanet, S. Hanneton, F. Bevilacqua. “Sensorimotor adaptation to a gesture-sound mapping perturbation”. *In Proceedings of the IXth International Conference on Progress in Motor Control (PMCIX).* Montreal, Canada. 2013

S. Hanneton, E. O. Boyer, V. Forma. “Influence of an error-related auditory feedback on the adaptation to a visuo-manual perturbation”. *In Proceedings of the IXth International Conference on Progress in Motor Control (PMCIX).* Montreal, Canada. 2013

Chapter 2

Basic concepts and literature results overview

Auditory feedback of movement learning finds echoes in many scientific domains, such as neuroscience, experimental psychology, sports, rehabilitation medicine or audio research. The fact that literature is scattered makes it less accessible and it may be difficult to compare results due to multiple goals and contexts among the studies. In this chapter, the first section describes basic sensorimotor control principles and the auditory-motor loop created in a gesture-sound interactive system. Then, a state of the art of the use of auditory feedback and sonification in motor tasks is proposed. We will browse through various techniques and applications in relationship with movement. We discuss first the large potential of movement sonification and second, the relative lack of results about learning in gestural interactive contexts. We will also describe historical fields of research in auditory-motor integration, that represent important and inspiring work and finally present important fields of application.

2.1 Principles and definitions about sensorimotor learning

In this section, we introduce learning theories, concepts and main results in sensorimotor control. We will expose basic concepts, that are widely used today, and sometimes referred to in our work. We first define the cognitive and sensorimotor levels as generally considered in neuroscience. We will then briefly summarize the main concepts regarding sensory feedback, and depict principles of sensorimotor control and learning, with a particular focus on basic notions on adaptive processes. Finally, we will give an overview of the main theories of perception-action coupling, illustrating the various approaches.

2.1.1 Sensorimotor and cognitive levels

During the interaction with an object, or while using a gestural interactive system, the user is involved in a loop where several levels of processes are involved: perception, information processing, decision making, action, etc.). A key feature of the physiological and functional characteristics of the biological sensors, actuators and of the nervous system involved, is the *dynamic* aspect of the interaction. Even in the case of voluntary actions, the system has very short time to take decision and perform actions. Brain structures responsible for planing, problem solving, decision making, cannot be activated continuously. Consequently, the separation between two levels of neural activity has been introduced in neuroscience, the sensorimotor level and the cognitive level (Paillard, 1985).

This distinction is partly based on functional roles of the neural processes. Kerlirzin et al. (Kerlirzin et al., 2009) define cognition as the mental processes that include acquiring, storing, transforming and using knowledge and expertise. They give the examples of perception, memorization, reasoning, information processing, problem solving and decision making as such processes. These cognitive processes are largely voluntary and conscious. The sensorimotor level refers to more low-level processes involved in transmitting to the brain signals coming from the sensory receptors, in the adjustment of muscle length and forces, and in the regulation of feedback loops involved in the control of movement. However, the two levels are not independent and have to collaborate during action-perception processes. As we explained in chapter 1, we focused our work on the sensorimotor system, and particularly on the function and adaptive properties of the sensorimotor loop (Wolpert and Ghahramani, 2000) regarding auditory inputs.

2.1.2 Sensory feedback

Feedback is defined by Hartveld (Hartveld and Hegarty, 1996) as an “information received back by the appropriate control centers in the brain on the resultant movement”. This sensory (somatic) information is provided during the production of a movement, or after a task achievement. Any movement of the body generates *intrinsic* feedback, which is a sensory information resulting from the deformations of the sensory organs (skin, muscles and tendons receptors) during the movement itself. It can be for instance tactile, proprioceptive or kinesthetic. This one word encompasses a very wide range of perceptive variables. Proprioceptive feedback for instance is always active while we move and is omnipresent. The vestibular system gives the nervous system information that can be qualified as intrinsic as it delivers data from inertial and inclination sensors in the inner ear. *Exteroceptive* feedback defines information from the outcome of the movement through the senses, like the vision of the body moving. This information can come from an observer, like a coach or a physical therapist. Hearing the auditory outcome of our actions can be considered as exteroceptive feedback, either its related directly to our body (sound of steps) or to the object or system we manipulate (from opening a lock to driving a car). This type of information, coming from outside of the body is called *extrinsic* feedback and can supplement intrinsic feedback taking many forms (auditory, visual, verbal, etc.).

Augmented feedback

Extrinsic feedback can be artificially created. In this case, it is called *augmented* feedback. It can take the same forms (for instance a visual cue or a verbal advice) but is generated by a specific equipment like computer, body sensors, etc). It can be conveyed through any sensory modality. Feedback is often named after the source of information it gives: 'biofeedback' refers to providing quantitative biological information that are normally non accessible, like the expression of a muscle force or electrical activity. Augmented feedback is implemented to add supplementary information to an active sensory channel: for instance, it augments the visual channel when providing supplementary written or iconic information on a display when performing a task.

Motion capture and sensing technology allow to provide concurrent feedback while performing an action. The system usually acquires information from the action with sensors and computes and deliver an 'online' feedback synchronized with the action. It can still be discrete or continuous information, and represents either body-related or task-related information.

Task-related feedback

In motor learning experiments, two types of feedback are generally delivered (Hartveld and Hegarty, 1996; Robertson et al., 2009), Knowledge of Performance (hereafter named 'KP') and Knowledge of Results (named 'KR'). They designate the type of message they deliver relatively to the context of the task that the subject is performing. A KP feedback gives quantitative or qualitative information about *how* the movement is performed, like a joint extension, or force profile. KR concerns the goal to achieve and tells about the subject's success (or failure) in achieving the task, like a confirmation or a success rate; the subject can use this information to perform better on the next attempt of his task. It is always related to an error, it implies also to define clearly the task, the reference performance and a measure of the distance between the action performed and the reference. This reference has to be easily accessible for the subject. For example, in a reaching task, KR could indicate whether the target was hit or not, and KP would inform about movement quality related to the goal, such as "move your trunk less" (Subramanian et al., 2010). KP and KR are usually terminal, meaning they are delivered *after* the task or after a movement. The term 'contigent' feedback, sometimes encountered in the literature, means that the feedback strongly corresponds with task behavior which can be controlled by the user. It can be for instance, information on velocity, effort, force, etc... It is usually provided during the task.

All in all, the classifications of feedback types depend on the point of view taken to consider the interaction, the task itself, the sensory modalities involved and the point of view of the user while doing the task. We will see that the classical separation of KP and KR types of feedback may not always apply to the concepts of auditory-motor interaction we develop in the present work. The specific case of using the auditory channel as a feedback media is treated in section 2.2.

Sensory feedback is permanently integrated by the CNS for retro-action and close-loop control. It is an essential component of attention and decision making, and is often integrated for multimodal data fusion, taking advantage of various information sources. The sensorimotor system is responsible for this integration enabling motor control and learning.

2.1.3 Sensorimotor learning and integration

Sensorimotor learning is obviously a time-related process in a dynamical environment. We review here the main concepts used in the discipline to describe the important features of sensorimotor learning. We emphasize that we only describe here principles of research fields related to our work.

The expression “sensorimotor integration” is quite wide and is used in many ways in the scientific literature. The use of this term seems to be case-specific and to depend on the modality concerned (vision, speech, action). Similarly to the notions of feedback, the various experimental contexts (developmental, skill acquisition, rehabilitation) serve various interpretation and concepts of the term, from a synonym of learning a specific task, to acquiring perceptual information and multi-sensory processing mechanism.

The key principle that allows for sensorimotor learning and adaptation is the capacity of the central nervous system to modify its own organization and functional network, particularly in the sensory and motor areas of the brain. This capacity is called cerebral plasticity or neuroplasticity (Pascual-Leone et al., 2005; Zatorre et al., 2012). It designates the ability of the neurons to reorganize their network in a purpose of optimization. Particularly, it allows to express new behavioral or perceptual features that are a better response to the environment or to the context. In a sense, sensorimotor learning is generally considered as an optimization process. Neuroplasticity is now accepted as the main substrate for motor sensorimotor learning, rehabilitation processes (*i.e.* re-learning) (Whitall et al., 2011; Masiero and Carraro, 2013) and sensory substitution among others (Bach-y Rita, 1970; Proulx, 2010) - see chapter 4. The neural mechanisms enabling this faculty has been firstly investigated by Hebb in the now called Hebbian theory (Hebb, 1949). It basically explains that the more a synaptic path is excited, the more likely the cells in this path see their efficiency to excite the following cells increase. This initial theory was later confirmed with the discovery of similar mechanisms called long-term potentiation (LTP) and long-term depression (LTD) (Squire and Kandel, 2000).

Prediction for control

It appears that anticipation and prediction are necessary conditions to produce adequate movement in a particular context; anticipation can be useful to compensate for processing and transmission delays of the nervous signals by the CNS. The aim being to build estimations of the future state of the body and the perceptual system. The idea of mental images enabling these predictions of the movements is generally accepted. Whether they involve semantic or analogical (graphic) information is still under debate (Kerlirzin et al., 2009). From a computational point of view, much research tried to formalize the prediction mechanisms, mostly for arm movements, pointing to a target under constraints, grasping objects, and walking actions. A popular prediction and control model has been developed by Wolpert, Ghahramani, Jordan and Flanagan (Wolpert and Flanagan, 2001; Wolpert and Ghahramani, 2000; Wolpert et al., 2005). Two types of prediction are distinguished, established from the current state of the body and its sensory input. A prediction on the future body state is made using a forward model of the body dynamics, according to the current estimated motor command that has been previously given (through an efference copy, its mental image). A sensory prediction is also constructed to estimate the expected sensory inputs consecutive to the movement execution and modification of the environment. When the sensory feedback is received (with delay, noise etc.) it is compared to the sensory prediction and serves as a correction signal to the estimation that is built. This model is often referred to as mixed forward and inverse model (Wolpert and Kawato, 1998; Franklin and Wolpert, 2011).

To be as extensive as possible, we cite here some other models of movement control and optimization, based on physical or statistical quantities optimization of movement features: the minimum jerk model (Flash and Hogan, 1985), the minimal effort model (Hasan, 1986), the minimum torque change model (Uno et al., 1989), the minimal variance model (Harris and Wolpert, 1998) and the optimal feedback control (Todorov, 2004). These models serve for instance to evaluate the ‘two-thirds power law’, which establishes empirically that the instantaneous speed of a continuous drawing movement is proportional to the curvature radius of the drawn trajectory to the power 1/3 (Viviani and Flash, 1995) in the case of a drawing movements[1].

The motor and perceptual predictions necessary for motor control are built and modified with experience over time. Various concepts of learning have been developed to account for the establishment of such models, describing how they are modified with experience and participate in acquiring new skills.

---

[1] It describes the way the tip of the pencil will slow down at the extremity of an ellipse and speed-up along the longer sides.

Learning

In (Franklin and Wolpert, 2011), Franklin and Wolpert detail three main types of learning in the field of sensorimotor control: supervised learning, reinforcement learning and unsupervised learning. Supervised learning is driven by the error between the target of the action and the action. This error is provided by an external agent named the supervisor. Reinforcement learning is driven by a positive or negative reward signals. Unsupervised learning is driven by signals not related to an error but by the experience of the task by the agent. Risks, rewards and costs (and not only error) are also involved in adaptation and could explain unusual gesture trajectories performed under specific circumstances such as sensory or motor perturbation, as can be seen in (Forma et al., 2011) and (Kagerer and Contreras-Vidal, 2009) or (Davidson and Wolpert, 2004).

As described by Subramanian et al. (Subramanian et al., 2010), *explicit learning* is the attainment of accessible declarative knowledge of components of a motor action through cognitive process that can be tested by component recall or recognition. The main differences with *implicit learning* are the awareness and consciousness aspects of the explicit learning. The latter is a form of consciously accessible knowledge.

Implicit learning defines an unconscious mechanism of learning involved in everyday skills and acquired through physical practice without the requirement of declarative instructions. A common illustration of this notion is learning to ride a bike, which is learned as a global action (evaluated in the result), without always thinking of all the movements involved simultaneously (keeping balance, steering, pedaling). In laboratories, implicit learning for discrete tasks is usually measured with serial reaction time task (SRT) tests. During SRT tests, participants are asked to respond to a set of stimuli, each of them requiring a particular action. Transition probabilities between some stimuli are controlled and adjusted so that participants can implicitly learn the transition probabilities. Assessing implicit leaning on a continuous task is more complicated and depends on the task itself and how the performance is defined (see (Lang, 2011) for framework and experimental approaches).

Contrary to learning, adaptation processes can be defined as a trial-by-trial modification of a movement, guided by the error feedback received during the task. It is a shorter term mechanism (minutes scale), defined according to Bastian (Bastian, 2010) by three criteria: a-the movement is a specific and identifiable gesture that needs to be changed in a parameter, b-change is gradual over a relatively short period of time, and c-adaptation is subject to a rise time and to 'after-effect' when the target is the normal gesture again (a de-adaptation is needed).

It has been proposed that adaptation could be achieved with two learning processes, one faster and one slower (Smith et al., 2006). These processes learn and forget faster and slower, respectively. This mechanism allows quick assimilation and better long term stability to perturbation and during de-adaptation process. Note that other models describe a unique fast process and several slow processes (Lee and Schweighofer, 2009). Franklin and

Wolpert (Wolpert et al., 2011) give many examples and conclude by saying that learning is optimally performed by the sensorimotor control system to integrate not only nonlinearity, non-stationarity and delays but also noise and uncertainty. They recently proposed that the optimization of movement during the learning is guided by an optimization constraint: to minimize the uncertainties. In addition, Franklin and Woplert (Franklin and Wolpert, 2011) wrote an extensive review and synthetic paper on sensorimotor learning and possible learning algorithms that could be used by the neural networks of the central nervous system and particularly by the cerebellum.

Brashers-Krug (Brashers-Krug et al., 1995) focused on plasticity and adaptation of the sensorimotor system. He studied de-adaptation and catastrophic interference during learning and suggests that motor learning can undergo a process of consolidation. Davidson (Davidson and Wolpert, 2004) confirms that learning can take a very long time (hundreds of gesture repetition) and de-adaptation is very often faster. Davidson experiments different scaling down processes during de-adaptation.

Changes in performance do not imply learning: as van Vliet writes (van Vliet and Wulf, 2006) the effect of feedback on motor learning often mixes the result of performance and learning. In a movement trajectory task, learning can be mistaken with simple adaptation. Maulucci states that (Maulucci and Eckhouse, 2001) adaptation denotes trajectory modification that takes place immediately after training *i.e.* at the end of the training session; learning denotes trajectory modification that takes place some time after training *i.e.* the retention and use of trajectory modification from one session to another. Studies rarely assess this point and evaluate usually short term effects. Long term studies or retention time tests still have to be carried out. Engardt (Engardt, 1994) raises also the question establishing a deep motor scheme during learning versus motor performance that can be easily washed out.

While studying how visuo-motor perturbations can influence auditory-motor representation, Kagerer (Kagerer and Contreras-Vidal, 2009) noticed that movement planning can benefit from previous visuo-motor transformation in the same limb movement. He suggests that new visuo-motor internal models developed after a perturbation can be used by the auditory-motor system: “the internal model formed during exposure to the visuo-motor distortion is immediately available to auditory-motor networks. In other words, speaking on a modeling level, the transformation of the difference vector between the visual target and hand position into the desired hand/joint kinematics and dynamics is being used by the system when the task suddenly becomes auditory-motor in its nature”.

Ostry and Darainy (Ostry et al., 2010) investigated how plasticity and structural changes also occur in the sensory areas of the brain. They showed a link between perceptual changes and motor learning: the perception of a limb can be altered after a motor learning task involving this limb. Interestingly, this effect did not appear when the task was done passively (with the experimenter moving the subject’s limb). This work confirms that motor learning has consequences in motor and sensory areas of the brain. They developed the idea of perceptual learning (Darainy et al., 2013) which describes the modification of expected sensory targets along with a motor learning process. They showed that this mechanism could support motor learning. Again, the importance of the action and active decisions is underlined.

The sensorimotor learning principles described previously emphasize the coupling between action and perception, and its importance in motor control. This coupling has been formalized at a more global level, with various visions which have evolved substantially in psychology and neuroscience.

2.1.4 Main theories about action-perception coupling

Many theories allowing to interpret the coupling between perception and action, and its role in motor control, have been established since the advent of cognitive sciences. We describe here the most important of them and precise the specificities of the auditory coupling that we consider. The large majority of the studies on action and perception have been carried out in the fields of vision and language development. More recently, the specific case of expert musicians served as a particular context of complex auditory-motor interaction. The perception-action coupling has been less investigated in the case of interactive sonic systems. We draw the reader's attention on some useful concepts that still need to be extended to interactive sonic systems.

Cognitivist approach

From a cognitivist point of view, it is essential to constantly process data coming from our sensory system when performing a motor action. Perception enables us to modify and correct a trajectory (either on a single limb or a global trajectory as while driving a car). The trajectory must be predicted, simulated internally and evaluated, to achieve the best performance. This requires the notions of motor programs and internal representation in the CNS (central nervous system), which is considered as purely an information processing unit. A typical example is given by Kerlirzin et al. (Kerlirzin et al., 2009): a waiter when lifting a bottle from his tray will automatically adjust the force in his arm as the bottle leaves its surface. On the other hand, if the bottle is removed by anyone else, it is impossible for the waiter to anticipate and adapt. Even if he receives warning he will not keep the tray still. Although this approach has historically served many experimental work and models (Wolpert and Flanagan, 2001; Wolpert et al., 2005) for motor control and prediction, others approaches question the capability of the central nervous system to build these internal representations of movements, especially 1- because of the quantity of information and schemes that would need to be stored and processed during the action, and 2- because the metaphor of the CNS as an information processing and storage machine can be challenged (Varela, 1989).

Enaction and embodiment

Varela, Thompson and Rosch (Varela et al., 1991; Varela, 1989) suggested an alternative to the cognitivist approach of a pre-determined world that the brain would need to learn and retrieve. They proposed that cognition should be considered as a situated action. Two main points characterize the notion of enaction: perception is seen as an action, and cognitive structures arise from sensorimotor schemes (which guide action with perception) (Varela et al., 1991). The theory of enaction explores the way a subject, doted with perceptual abilities, can guide his actions within a local context. As the context is constantly changing, due to perceptual action from the subject, perception is not deciphering the pre-determined world, but determining the sensorimotor relationships describing how action can be guided. The mind and the body being functionally connected in both directions represents the *embodiment* of the human experience. This notion tends to blur the theoretical separation of cognitive and sensorimotor levels. It is also close to (and inspired by) the phenomenological approach by Merleau-Ponty (Merleau-Ponty, 1945). The concept of embodiment has been recently applied to many fields where interaction is central, like music cognition and production (Leman, 2008), communication, human-computer interaction, and gesture-sound interaction (Caramiaux, 2011). For an application of enactment and 're-enactment' in sonic interactions see (Schnell, 2013). Implementing and testing models derived from the notion of enaction is still a challenging question, although some example with musical or sonic interactive systems can be found (see section 2.2).

Ecological and direct perception approaches

For Gibson, no pre-existing model or representation of the outer world is needed to actually perceive; in the ecological approach, the information is 'given' to us through the stimulation of the perceptual system (Gibson, 1986). The foundation for perception is ambient, ecologically available information. This theory aims at re-locating the subject in his environment, hence the term of 'eco-logical' or ecological. In the framework of the ecological theory, Gibson introduces the concept of "affordance", which represents the possible actions that an object offers in a particular context or environment. These affordances can be interpreted as actions which are directly perceived and differ from the actual physical properties of the object.

Inspired by an ecological thinking, the general τ (tau) theory is an example of these principles. The body is considered as a whole entity, interacting with its environment from which it gets a continuous flow of information. The general tau theory, also referred to as theory of perceptual guidance, formalizes movement control as the closure of a spatial and/or force gap between the effector and the goal of the movement (Lee et al., 1999). This closure is ensured by the regulation of the taus of the gap, which are the time-to-closure of the gap at the current closure-rate. Under this framework, movement coordination is modeled as keeping the taus of the gaps involved in a constant ratio, called τ coupling (Lee

et al., 2001). This theory states that sensorimotor control is achieved by constantly sensing the taus of the gaps and keeping them coupled. It has been observed in both human (*e.g.* in sports or daily activities actions) and animals (like in bats using echolocation).
Also inspired by the ecological theory, the direct approach assumes that the central nervous system does not have to make calculus or computation in order to draw a link between sensations and actions. It only has to find in the environment the appropriate signals that have to be properly associated with the correct motor response. Experimental results show that observation is linked to imagined actions to predict forthcoming action effects, and tend to demonstrate the importance of self-perception to predict actions outcomes (Knoblich and Flach, 2001).

Motor theory of perception

In this framework, perception is an action that is simulated. Efficient adaptive models have been proposed combining feedforward and feedback controllers (Wolpert, 1997; Wolpert and Kawato, 1998; Miall and Wolpert, 1996). Motor theory of perception considers the CNS as a simulator that uses internal models and predicts future perceptual states (resulting from the application of motor commands) in order to take decisions. Berthoz (Berthoz, 2003) formulated that the experience of the body (and memory) allows the CNS to anticipate the sensory and motor consequences of the action.

Sensorimotor theory of perception

A similar approach, the sensorimotor one, draws a link between motor actions and feelings. In (O'Regan and Noë, 2002), O'Regan and Noë stand that we don't need to build a representation of our environment in our brain to experience the outside world[2]. The relationships between the motor commands and their sensory consequences respect sensorimotor 'contingencies' which originate from the structure of the body. The feeling of seeing, for instance, is not generated by the activation of an internal representation but by our active information inquiry that we can access cognitively. The commitment in a sensorimotor action leads to 'active' perceptual sensations; this notion of commitment and awareness is crucial, in order to know that some movements can have consequences on our sensory inputs. Numerous examples of application of this principle are given in (O'Regan and Noë, 2002), all referring to vision.

A recent paper by Engel et al. (Engel et al., 2013) emphasizes the enactive, interactive and sensorimotor conceptions of cognition. The authors argue that cognition should be seen as supporting action and based on sensorimotor coupling. Experimental findings supporting this paradigm in visual and auditory domains (Aytekin et al., 2008). The interested reader can read the extensive essay by Anderson on embodied cognition (Anderson, 2003). This

---

[2] This theory has been developed in the case of visual perception.

view tends to blur the dichotomy which has been established between cognitive and sensorimotor levels of neural activity.

The debate on the cognitive versus sensorimotor levels and the different visions of action-perception coupling is still open. A common tendency is though to consider the importance of action in behavioral studies. It is likely that new interactive scenarios (constantly motivated by technological development) will bring future perspectives on sensorimotor learning. The auditory modality, which still suffers from a lack of interest for exploring sensorimotor mechanisms, already showed promising results as a feedback channel for sensorimotor learning.

2.2 State of the Art: auditory feedback of motion

The large majority of neuroscience research addressing the human sensorimotor system deals with visual, haptic and vestibular sensory inputs, and rarely mention the auditory modality. Historically, most papers that report on auditory-motor mechanisms addressed speech learning and production. More recently, specific studies on musical experts explored the audiroty-motor interactions in music playing and learning. On the one hand, due to promising applications in movement learning (mostly in sport and rehabilitation fields), there is an increasing number of studies showing the potential interest of auditory feedback, although the technology used in these types of applications remains generally rudimentary and the interaction quite poor. On the other hand, the most advanced interactive gesture-sound systems are found in the sound and music computing field, but there is an important lack of evaluation of these systems and very few studies focused on sensorimotor learning.

Evaluation of performance and learning in a gesture/sound interactive system is obviously linked to the context: mechanisms and results are certainly different within a musical control context or in motor rehabilitation of a stroke survivor patient. As previously mentioned, the use of movement sonification is quite recent and scattered over different research fields. That could partly explain why there is a lack of coherence between the results available. Common basis and formalisms are still needed in this domain, especially regarding continuous interactions. The various fields bring their own angles and contexts, whereas context is thought central in the perception-action loop according to neuroscience (see section 2.1). In this section, we intend to introduce the concept of sound-movement coupling as it is studied today in the literature and to overview the current main application fields of auditory feedback and movement sonification.

2.2.1 Auditory feedback and sonification

Auditory feedback globally designates the provision of information during a task through the auditory modality. This term is used regardless of the substance (the meaning, like KP or KR) or the form (acoustic features, continuous or discrete events). Of course, many actions naturally provide auditory feedback, caused by physical interaction (with materials, between objects, etc.). What we mean by ‘auditory feedback’ is the provision of a supplementary information, or the augmentation of an available information using audio signals. The referenced techniques are *audification*, *auditory icons*, *earcons*, and *sonification*. They are often referred to as *auditory display* when these concepts are used to translate an information into an auditory signal carrying a metaphor or a meaning in a particular context (Walker and Kramer, 2004; Walker and Kramer, 2005). These techniques are well defined in (Hermann et al., 2011). Audification, the most straightforward technique, directly translates the waveform of a periodic data into sound. Minor signal processing is often operated, such as resampling or frequency shift. A recent interstellar example has been provided by the European Space Agency with the ‘sound’ of the Philae lander touchdown on the comet 67P/Churyumov-Gerasimenko[3]. Auditory icons are the sonic equivalent of a visual icon (like on a computer system). They are short non-speech sound samples, imitating ‘real’ sounds. The message is conveyed by the metaphorical evoking content (ecological approach (Gaver, 1993b)). Hermann et al. (Hermann et al., 2011) defined earcons as “short, structured musical messages, where different musical properties of sound are associated with different parameters of the data”. For instance, an ascending melody would be used to indicate an upward action confirmation. Extensive research has been carried out on these auditory displays, conceptualizing their design and utilization, and gathered in the field of Sonic Interaction Design (SID) (Franinović and Serafin, 2012a; Franinović and Serafin, 2012b) (see 2.3.1).

Sonification, in a general way, allows for richer sound generation and more complex interactions through the auditory modality as it can take many forms (discrete, continuous, direct, time-related, metaphorical, etc...). The resulting acoustic signal translates data relations and magnitudes into perceivable acoustic dimensions. Many examples of systems, techniques and methods can be found in (Hermann et al., 2011) and (Dubus, 2013). Biological signals can also be sonified like EEG for clinical and research applications (Baier et al., 2007) or EMG (Pauletto and Hunt, 2009). Novel computational and mapping techniques offer wide possibilities, including physical model synthesis; see for instance (Françoise et al., 2013b; Françoise et al., 2013a; Schwarz, 2011). For about the past ten years, a scientific community has built up around the use of sonification for various usage (music technology, art or sport sciences). The constantly developing technology gradually broadens the spectrum of applications, but the research studies do not always keep up[4].

In the following section, we intend to overview the range of applications[5] of sonification for motor learning that has been published in the scientific community. Several large review articles are referred to, describing experimental results or methodology points. However, very few papers tend to generalize principles of auditory feedback and sonification design

---

[3] `blogs.esa.int/rosetta/2014/11/20/the-sound-of-touchdown`

[4] In 1999, Kramer et al. pointed that much work remained to be done with sonification (Kramer et al., 1999).

[5] Rather than techniques.

to enhance motor control and learning. Definitions and basic concepts are rarely consistent over the authors.

The work developed in the rest of this document will focus on the use of continuous sonification of movement features and its relation to motor learning in various experimental context.

2.2.2 Sound-movement relationship

It seems quite straightforward to think that sound has a natural influence on our movements, as any motion involving physical contact would produce a sound. An interesting link between sonic features and gestures has been drawn by Caramiaux et al. (Caramiaux et al., 2014). They showed that if the cause of a sound can be identified, a subject would perform a spontaneous gesture that tries to mimic the action that probably produced the sound. If the sound contains no perceivable causality, the movement would trace acoustic features profiles, with less inter-individual variability. This suggests that acoustic features of a sound can refer to a metaphoric gesture with the sound being the reference among the subjects. Merer et al. (Merer et al., 2013) found a topology of motion features evoked by sound (assessed by drawing after hearing a sound). They distinguished large categories such as direction (rising, descending, and horizontal), speed variations, oscillations or shape. A study on the rowing movement is presented in (Schmitz and Effenberg, 2012), where resulting sounds from movement sonification of two kinematic and two dynamic variables were presented to subjects (including their own gesture in the corpus). The resulting sounds proved to be rich enough to allow for prediction and identification of the movements, leading the authors to conclude that sonification of motion features can enhance perception and motor prediction.

The link between sound and motion has also been illustrated using functional MRI measurements. In (Schmitz et al., 2013) the authors showed to observers video sequences of an avatar performing swimming movements, along with congruent or incongruent sonified velocity components. Congruent sounds led to a better perception of motion and to increased activity of the action observation areas in the brain. With similar stimuli, Scheef et al. (Scheef et al., 2009) showed the importance of a specific region in the brain (V5/MT) for multisensory processing of motion. This area was activated in the presence of sonification of a motion displayed on video. According to the authors, this might explain how sound emphasizes visual motion and contributes to motor learning as an extrinsic feedback.

Similarly to our approach from a system point of view, Vinken et al. (Vinken et al., 2013) and Brock et al. (Brock et al., 2012) proposed the use of inertial measurement units (IMU) for movement acquisition and sonification. They used MIDI mappings between body segments angles, and MIDI notes in (Brock et al., 2012), and continuous frequency/amplitude/spectral/panning features mapped to movement components in the anatomical planes of the body in (Vinken et al., 2013). In this last study, the authors investigated the potential information content of the sound produced when doing everyday actions, like pouring liquid into a glass, drinking, brushing ones teeth, etc. Each action were described by a verb, but they always referred implicitly to an object to interact with. The authors asked participants to discriminate between six action sounds produced by the sonification of each action. They found that the participants can discriminate between actions, and compared the efficacy of movement recognition based on continuous sonification with visual identification under point-light display conditions (*i.e.* seeing only points on body joints, similarly to motion capture systems like the Kinect device). They noticed that even low-dimensional sonification allowed for the identification. Their results were actually dependent on action structures and categories. They related their results to auditory-motor coupling, supporting the idea to use sonification for movement training or rehabilitation (in multimodal conditions).

‘Positive’ and ‘negative’ aspects of music and speech have been proposed in (Vogt et al., 2009). The authors showed that positive sounds (in favor of the targeted movement) helped the participants on comprehension and motivation whereas the negatively designed sounds were inhibitory. The design of the sounds was very much metaphorical (sound textures, voices) and evolved during the session (to avoid repetition). Participants could also play with their preferred music. Results suggest that movement sonification can improve concentration and attention when learning new skills, although it was only a pilot study.

Kinetics

Among the almost endless design solutions available for sonification, Effenberg (Effenberg, 2004; Effenberg et al., 2005) evokes the ecological approach of perception in kinesiology, where a close relationship between movements kinetics and sound can exist. He describes that many kinetics events are *per se* related to the production of a sound. The motor system uses these sounds as an important input parameter - this has also been shown in sport (Takeuchi, 1993) (see below). This relationship could be applied to silent part of human gesture to bring more information about the movement, especially the velocity. This paradigm is thought to be beneficial for motor control and motor learning according to him. He argues that functionality and origin of sound need to be accessible when using sonification to get proper extrinsic feedback. He also supports that spatio-temporal features have to be respected regarded movement sonification to activate multisensory response of the system.

De Götzen et al. (de Götzen et al., 2006) focused on Fitts’ law to sonify gesture parameters. Fitts proposed in 1954 (Fitts, 1954) a mathematical model of human motion allowing to estimate the movement time in a reaching task depending on the width and distance of the target. It is nowadays an ISO standard to evaluate pointing devices; it explains indeed the speed-accuracy trade-off of the reaching movement. In their experiment, de Götzen et al. proposed a tuning task while hearing a pure tone feedback. Participants had to reach a target frequency under audio, visual and multimodal conditions. Auditory feedback on the movement seemed to help participants with difficult tasks when presented along with visual feedback. Accuracy was better but speed decreased with auditory feedback. These findings can be related to musical control. The authors also noticed that using an auditory feedback alone, without vision, was difficult.

Trajectory control

In the framework of the closure gap and Tau theories, Rodger and Craig (Rodger and Craig, 2011) investigated the synchronization of tapping (left to right finger and hand movements) to continuous, dynamic or discrete auditory beats. With discrete sounds the synchronization error was smaller but the variability was lower with continuous sounds. Continuous feedback also proved to give rise to more sinusoidal finger movements and less jerky trajectories. This illustrates the potential of continuous interaction for motion control. Several studies emphasized the potential of multisensory (including auditory) temporal feedback for movement synchronization, like (Wing et al., 2010).

Castiello et al. (Castiello et al., 2010) went beyond visual perception of hand movement by measuring the effect of sound cues on visually-guided movements. They showed that the presence of an auditory stimuli (recorded sounds of materials) that is congruent with the visual information available can modify participants end point movement. Participants tended to reach and grasp an object (covered with two materials) by the part covered with the material heard in a sound presented before or after the initiation of the movement. This emphasizes the role played by auditory input in the representation of objects and their affordances, as well as the multi-modal aspect of neural action mechanisms.

Drawing and writing

Hand writing have been also studied, as a case of gesture/sound interaction with high prior training. Thoret et al. (Thoret et al., 2012; Thoret et al., 2014) studied drawings sonification to investigate relationships between gesture-produced sounds and gesture recognition. They focused on the speed of the gesture used as the input of the physically-based friction sound model. Subjects were able to recognize gesture trajectories with the friction sound they produced or generated by the model, if the sound mappings respected the two third power law. They insisted besides on the relevance of timber variations for the perception of movements. In (Kim et al., 2012), the number of -Chinese- characters written per minute was increased when subjects received an augmentation of the sound of writing, picked up on the paper with a piezo cell. Danna et al. developed a sonification of writing based on velocity peaks of motion, dedicated to diagnose and treat dysgraphia in young children with the help of an auditory feedback (Danna et al., 2013b; Danna et al., 2013a; Danna et al., 2014). They obtained encouraging results, emphasizing the use of velocity data for sonification but also motivational aspects which are crucial with children.

As it can be noticed, many papers reported on auditory perception of movement-generated sounds or sonification, and only a few investigated the dynamic and continuous *coupling* between sound and gesture. The next section reports studies with results on sensorimotor assessment.

2.2.3 Learning (with) auditory feedback

Motor learning of temporal structures of a movement can be improved with auditory feedback (van Vugt, 2013). On the one hand, results showed that non-musicians could benefit from temporal information provided by auditory feedback, and that they are sensitive to distortion of that temporal feedback. On the other hand, studies with expert musicians (see section 2.3.2 for more details) surprisingly revealed that they have become quite independent of the auditory feedback.

A fair number of studies applied ecologically-inspired mappings between sound and movement; it is the case for most of the papers cited below. Rath et al. (Rath and Schleicher, 2008) showed that an ecological mapping could fasten learning compared to a more abstract auditory feedback. Although, in the particular case of this study, the abstract sound led to a generally higher level of performance, even if it was less appreciated by participants. The authors also pointed out the relevance of a sonification based on the velocity of an object - here a virtual ball balanced with a stick.

Reference movement

An example of continuous movement sonification for skill learning has been developed by Sigrist (Sigrist et al., 2014) (see also (Sigrist et al., 2011)) in the case of the rowing task, which is considered quite complex, due to numerous degrees of freedom to control trunk and arms, sizeable amount of strength required, but also precision. The task actually consisted in learning specific oar angle and velocity profiles while receiving either auditory-visual or visuo-haptic feedback on a virtual reality-based rowing simulator. The sonification was based on the oar horizontal angle which controlled the pitch of a violin sound. Subjects received in one channel of the stereo headphones their own movement sonification, and in the other, the sound of the reference movement being sonified the same way. They conducted the experiment during three days, and feedback was present during only half of the trials. The results showed that participants adapted more efficiently their velocity profiles with auditory-visual feedback than with visuo-haptic feedback. Also, performance was less sensitive to the absence of feedback in that case. Interestingly, the spatial error sonification (angle) did not lead to significant improvements.
In this study, subjects had to extract information in the feedback from the comparison

of their motion to a reference profile. In some cases, this reference can be identified. In others, the reference may be either non existent as a unique solution or difficult to define properly from a kinematic point of view. Common features to any possible movement can generally be defined, but complete detailed gestural pattern may not be unique or accessible, especially due to necessary non-stopping adaptation (like in table tennis for instance, where the fine tunning of your movement depends on the opponent strike; see appendix A). The rowing task has also been studied by Dubus et al. (Dubus, 2012; Dubus and Bresin, 2014) who proposed different sonification methods using mobile technology. They confirmed the ability of participants to acquire simple movement features from sonified data, but did not measure significant effects of sound on the performance.

Tracking and pointing

A common strategy to design auditory feedback for movement control has been to sonify the error between a correct movement and the one performed by the subject. In this case, auditory feedback is often an augmentation of the visual feedback. Oscari et al. investigated this strategy in a one dimensional tracking task (vertical displacement), with a motor perturbation and binaural auditory feedback (Oscari et al., 2012). The reaching movement was perturbed by an orthogonal viscous force and adaptation was assessed with and without auditory feedback. Their results showed similar adaptation and after effect magnitudes with and without the auditory feedback, but suggested that it can be integrated to modify internal models. Similarly, Rosati et al. (Rosati et al., 2012) compared auditory task-related and error-related feedback in a tracking task with horizontal target trajectory. The aim of the task was a tracking exercise with a minimum-jerk trajectory. They confirmed that auditory feedback can improve performance in a task where vision is required to execute it, and found that it could help learning to adapt to a perturbation. The authors also found that task-related feedback improved performance during a complex and unpredictable tracking task and that, relatively to the auditory modality, a constant-length task can be executed better than a variable-length task. Unfortunately, no control group was included in the protocol. The results seem to show that error-related feedback (sonification of the error) was not beneficial and even deteriorate adaptation to a visual perturbation.

Pointing has been studied with spatialization of sound in a virtual environment in (Forma et al., 2011). The authors presented a system based on magnetic hand tracking where a continuous real-time sound feedback indicated the relative position (or distance) to a virtual sound-emitting target. The particularity of this experiment stands in the protocol during which the feedback was presented as if the “ears” of the participants were on the tracked hand. In this case, where the effector (the hand) and the sensor (the virtual ears) are spatially coincident, precision was slightly better (less jerk, and more direct trajectories). Subjects were able to learn this system, but the authors pointed out the large inter-individual variability. The question of audio spatialization for pointing tasks is addressed in chapter 3.

A recent study by Schmitz and Bock (Schmitz and Bock, 2014) investigated adaptation processes in the classic task of pointing to visual or auditory sources. They compared the re-calibration (washing out aftereffects) with either direct or indirect and visual or auditory feedback about target position. The auditory feedback was indicating an angular deviation (an 'error') by shifts in the fundamental frequency of a whistle sound. The visual feedback was also based on angular error, and the indirect one, instead of indicating the direction of the target, used a color code for the deviation magnitude. The authors intended this way to have a visual equivalent of the 'naturally' indirect auditory feedback (the sound has to be decoded to access a spatial information). Targets were either visual -light dots- or auditory -small movable loudspeakers delivering a mix of pure tones- on six different azimuths. Movements were repeated quite quickly: 26 moves in 45 seconds. The adaptation was assessed by rotating the coding of feedback information. First, the authors observed that pointing to auditory target with auditory feedback led to the same precision as pointing to visual target with visual feedback. They also observed that adaptation was slower with the auditory feedback but led to the same amount of correction; they interpreted this finding by a higher cognitive cost to use the 'indirect' auditory feedback, which emphasizes the notion of direct versus indirect feedback the authors introduced. Willing to compare the effects of visual versus auditory feedback, they concluded that the modality itself seemed less important than the method of feedback delivery. Again, this result shows that context and task are critical to assess learning. Besides, this study is another example of the use of an error-related feedback on spatial position.

Guidance

The question of auditory information content is one of the important issue of providing efficient feedback. Van Vliet and Wulf (van Vliet and Wulf, 2006) states that "prescriptive" feedback is more efficient than "descriptive" feedback in healthy subjects for a motor task (this might be the case of KP versus KR). It appears more efficient to give both errors information *and* suggestions on how to correct motion rather than giving error feedback only. Although continuous sonification during a movement can be motivating and informative, van Vliet and Wulf suggested that presenting a movement-concurrent feedback can decrease performance once it is turned off. They pointed out that concurrent and instantaneous feedback can disrupt intrinsic error estimation on the movement and lead to dependency of the subject. This effect has been referred to in the literature as the guidance hypothesis (Salmoni et al., 1984). A solution can be to introduce a few seconds of delay before providing the feedback but this would prevent real-time interaction and sonification in our case. Some authors also suggested to reduce the rate of feedback presentation (Winstein and Schmidt, 1990). Van Vliet and Wulf also proposed to reduced the rate of feedback during trials - 50% of trials with no feedback has shown better learning - and to search for a compromise between summary feedback (about every trial in a set) and average feedback (averaged over a set of trials). Subjects themselves could also choose when to receive feedback, suggesting better concentration and information processing. Technically, feedback can also be delivered only within a range of performance ("bandwidth feedback") or when the performance is outside the bandwidth ("quantitative feedback") although this seems less effective, according to the authors. Alternately, Buchanan and Wang (Buchanan and Wang, 2012) suggested that guidance develops during learning "because the training context favors the development of a spatial or visual representation of the action". As a result, they argue that a feedback present all the time is not necessarily detrimental. In any case, this literature unfortunately did not address continuous auditory feedback on motion.

Ronsse et al. (Ronsse et al., 2011) investigated the neural aspects of augmented feedback on a bi-manual coordination task. The results given in this paper refer to the guidance hypothesis with auditory feedback, at a neural level[6]. Two groups of participants were trained to blindly rotate left and right wrists on a manipulandum with a 90°out-of-phase coordination under fMRI scan. One group received visual feedback of the coordination and the other received an auditory rhythmic feedback playing alternative pitches indicating the coordination rate. Both groups improved with training, slightly slower for the auditory group, but behavioral and fMRI results showed that the visual group became dependent on the feedback and presented neural signs of ongoing reliance after training. A promising result is the capacity of retention of the auditory group observed in this experiment, in spite of the relatively long training protocol (4 days in a row before the day of the post-test).

More references

A recent review by Sigrist et al. (Sigrist et al., 2013) examined numerous papers on multi-modal and unimodal feedback with healthy subjects (rehabilitation patients are thought to use and benefit in a very different manner from augmented feedback for motor learning[7]). This review focuses on classic motor learning concepts and presents results relatively to them (such as the guidance and the specificity of learning hypothesis). First, the authors remind the large contribution to motor theories and feedback optimization for motor learning brought by visual and visuo-motor studies. Numerous results are detailed about concurrent, frequent and terminal feedback, task complexity, retention and guidance effects or adaptive feedback (changing the feedback rate display or threshold in the case of bandwidth feedback). The notions of self-feedback, self-efficacy, self-estimation of error and motivations are also mentioned. Many examples of auditory-augmented motor tasks are depicted, in various domains and especially in sports and skill acquisition. The example of sport shows that natural (*i.e.* causal) auditory information are used by performers to progress, such as in table tennis where the sound of the bouncing ball can provide high level information about its spinning movement (see appendix A). Furthermore, as previously

---

[6] For the first time, to our knowledge.

[7] This topic is address in section 2.3.4 and in the perspectives of the thesis

mentioned, extra visual feedback may overload the cognitive process in tasks mainly based on visuo-motor actions.
Sigrist et al. depict auditory feedback in three categories: auditory alarms, sonification of movement variables and sonification of movement error for motor training. Auditory alarms are considered to be easy to interpret, but do not allow for a continuous representation of the movement and do not give information on possible corrections to make. Sonifying movement variables seems more motivating for the participants according to the literature, but the authors argue that non-experts could not benefit from the sonification as it does not indicate clearly the correct movement to acquire[8]. Sonifying the movement error could solve this issue, according to them. It is interesting to note that the authors considered this sonification strategy as an entire category, as the alarm sounds. The authors argue nonetheless that few studies investigated sport training with error sonification, probably because, as we pointed out previously, it is very difficult to determine a correct target (or reference) movement in any case. Simple strategies of movement variables sonification have been explored in the literature, but general studies about how to facilitate motor learning and evaluation of the design of such process are still missing. Error sonification proved to reduce movement error in various tasks, but the retention of such learning and the potential dependency of subjects on the additional feedback still need to be explored. Generally, auditory feedback proved to be beneficial in various cases with a context-dependent approach, but the authors deplore that no study explored auditory feedback regarding motor learning theories.

Dubus and Bresin reviewed 179 papers reporting on the use of sonification of physical quantities, in order to analyze methods of mapping, physical dimensions used and to detect successful and unsuccessful strategies (Dubus and Bresin, 2011; Dubus and Bresin, 2013; Dubus, 2013). The main finding they reveal is that sonification mappings are rarely properly evaluated in the literature. As far as the mappings are concerned, a large number of papers described direct mappings representing simple kinetics magnitudes with pitch or loudness. They noticed a tendency towards ecological perception of sounds (although with simple acoustic attributes) and the extensive use of pitch to convey dynamic information. The majority of the studies the authors reviewed used low-level and sample-based synthesis, which encourages to go towards more complex sonification. It also encourages us to focus on the evaluation of motor learning in interactive systems, taking into account the specific context conditions in the evaluation and inter-individual features.

2.3 Connected fields and applications

We describe in the following paragraphs some important research connected to our work. Our project has been partly inspired from research and prototypes developed in these fields: sonic interaction design, music production, digital musical instruments, sports

---

[8] A question we will address in this thesis

and rehabilitation. Sonic interaction design proposes use-case scenarios and interesting sonification strategies. Music cognition and production investigate deep auditory-motor coupling either from the angle of musical sequence perception or targeted musical features. The digital musical interfaces community brings a novel approach of mapping strategies; these last two fields are now embracing embodied interaction concepts. Sports and skill acquisition domains obviously focus on movement production but offer different contexts where timing, proprioception and body perception are central. Finally, the rehabilitation field (which will be discussed in the discussion section of this work, 8.5) focuses on movement production, perception and control as well, and appears for us as a key application field to consider auditory feedback and sonification.
Specific scenarios of gesture-sound interaction are depicted in this section and a number of important results are emphasized. Important reviews are also referred to, as we know this section not to be exhaustive.

2.3.1 Sonic Interaction Design

The field of Sonic Interaction Design (SID) aims at exploring sound as an “active medium” (Franinović and Serafin, 2012a) and its use in various disciplines (arts, technology, architecture, design) to create sonic experiences both at a perceptual and social levels. Among the numerous prototypes and applications, we mention a few typical examples of this field; many of them can be found in (Rochesso, 2011) and in (Hermann et al., 2011). Data visualization, pointers and navigation with sound has been quite investigated (Rochesso, 1997; Bouchara et al., 2010; Lorenz et al., 2013; Katz et al., 2007). A growing interest is shown to interactive sound design in virtual reality environments (Serafin and Serafin, 2004; Wellner and Zitzewitz, 2008; Sikström et al., 2014) (see also chapter 4). Sound design also aims at augmenting object manipulation with sound (Lemaitre et al., 2009; Lemaitre and Susini, 2009; Geronazzo et al., 2013). Continuous sonic interaction is proposed to serve as a media for better object manipulation or to create new usage. Proposed scenarios often include earcons and auditory icons, although some used continuous sonification.

The action of screwing a moka pot (which should be very precise) has been sonified using an elasto-plastic friction model[9] on the screwing force, with the metaphor of a violin sound (Rocchesso et al., 2009). The dynamical change of timber quality proved to help participants achieving the right tightness.
Home improvement works are strongly multimodal situations, where proprioceptive, tactile, visual, olfactory and auditory feedback are useful. In addition, the attention is quite focused (to avoid injuries) and it involves strong physical interactions with materials. A prototype of augmented musical spirit level has been proposed during the Legos project[10]. In (Grosshauser and Hermann, 2010) the authors proposed to add extra auditory information about the inclination of a drilling machine. Pulsing sounds were used to indicate

---

[9] SDT toolkit `http://soundobject.org/SDT/`

[10] `http://legos.ircam.fr/workshop-participatory-design/`

orthogonality to the drilling wall, mainly vertical or horizontal angles.

Dirti (Savary et al., 2013) and the Pebbelbox (O'Modhrain and Essl, 2004), are examples of tangible interaction where the interface is a representation of the audio synthesis engine. Here granular synthesis[11] is controlled by seeds and pebble scrambled in a box, imaging the 'grains' of sound. This kind of design goes towards an ecological (almost embodied) sonic interaction, blurring the mapping function between sound and movement with highly metaphorical sounds.

Experiments observing the perception of action sounds have also been developed in the SID field. Subjects were able to perceive the spinning velocity of a metaphoric wheel by hearing the sound of a rack (Lemaitre et al., 2009). In the next experiment, subjects were asked to activate a vertical pump controlling the virtual rack to achieve a targeted spin velocity. Results showed that this sonification helped participants in the task, although they were not aware of it.

Unfortunately, sensorimotor-learning is rarely assessed in the sonic interaction design field, which developed its own methods, design and evaluation procedures. Nevertheless, inspiring prototypes and sonification strategies can be found (see for instance (Rath and Schleicher, 2008) commented in section 2.2).

2.3.2 Music performance

Musical production and neuroscience studies

In the context of gestural control of music and music performance, the movements performed are closely related to the auditory system since each move produces or influences a related sound (often related to the physical parameters of the gesture). This close relationship is reflected as well in common behaviors such as tapping one's foot or fingers when listening to music. Zatorre et al. (Zatorre et al., 2007) emphasized the fact that music listening and playing represent a rich and complex paradigm to study auditory-motor interaction. Particularly, the expert musician's brain has been recognized as a remarkable case of study for cerebral plasticity (Münte et al., 2002; Zatorre, 2003).

The temporal aspect of this auditory-motor synchronization in music has been highlighted for instance in (Zatorre et al., 2007). The presence of metrical structure in sounds seems to be enough to elicit auditory-motor interaction. The question of how auditory information are integrated to the motor system is wide and still under investigation (Chen et al., 2009). After studies in primates, Warren et al. (Warren et al., 2005) proposed a model of auditory-motor transformations based on audio templates with which auditory input information are compared. Auditory representations emanate from this comparison and are used to constrain motor responses. The authors insist on the central role played

---

[11] Synthesis method based on resampling of short sound 'grains', with probabilistic controlled sequencing. A sort of 'quantum' sound generator, from pre-recorded samples.

by the dorsal auditory pathway in the internal link between the auditory representations and motor areas of the brain.

In this musical context, the motor system must achieve high level performances about timing, but also sequencing and spatial organization. Movement features of musicians are interdependent on their own sounds produced and on external sounds (other musicians or playback track) (Zatorre et al., 2007). In (Conde et al., 2012) the authors used SRT (serial reaction time) tests to investigate the role of task-irrelevant auditory feedback during motor performance in musicians. They showed that musicians could benefit from auditory feedback, by increasing right hand performance but learning abilities did not change.

An interesting effect of musical production has been studied by Mathias et al. (Mathias et al., 2014). They observed that auditory-motor interactions in production experiences enrich memory and support the idea of motor predictions formed in the brain. The experimental protocol included performing and listening to melodies under EEG measurements[12]. The roles of memory and mental images have also been studied in (Brown and Palmer, 2013). The authors showed that auditory images allow to integrate auditory feedback more robustly against perturbations (auditory or motor). The imagery process proved to have a different effect on encoding or retrieving of musical sequences. Particularly, a high imagery capability led to better temporal regularity in presence of auditory perturbation in pianists. The authors draw indeed a link between skilled musicians and athletes in terms of action planning and mental imagery. A review by Novembre and Keller (Novembre and Keller, 2014) brings evidence as well for a strong action-perception coupling in the musician's brain, and for predictions and mental representation of actions.

The embodied cognition approach[13] has been recently applied to music production and perception (Leman, 2008). The motor system seems to play a role in music perception through sensorimotor and emotional interaction. Maes et al. (Maes et al., 2014) recently browsed in their review various studies supporting the idea of the embodied aspect of music cognition. They showed evidence that motor dysfunctions can cause perceptual disabilities. They also recommend to integrate, not only perception and action, but also introspection and social interactions as parts of the dynamical process of music cognition[14]. In this framework, the role of voluntary actions and action-related perception is of importance. Actions on the musical instrument itself, generating tactile perception when the instrumentalist plays it, can indeed be used to improve the timing accuracy of a musical performance (Palmer et al., 2009).

Obviously, music perception and production studies largely address the role of rhythm and often use auditory-motor synchronization with musical rhythm as an experimental paradigm. Rhythm is a particular aspect of music in which auditory-motor interactions are

---

[12] From an experimental point of view, the music cognition and production field often mixes behavioral and cerebral activity measurements.

[13] See section 2.1.4

[14] The relationship to others is also developed in (Novembre and Keller, 2014)

extremely advanced; consider for example percussion players or tap dancers. Tapping to the beat can be defined as a feed-forward auditory-motor interaction: auditory information is predominant on motor output as prediction and expectations are made (Zatorre et al., 2007). Feedback interaction also allows for continuous control of the sound, and enables musical expressiveness. The spatio-temporal characteristics of rhythmic movements are also investigated using a technique called auditory delayed feedback. The idea is to provide a temporal auditory cue on a movement or a sequence and to introduce a delay of presentation. Under this experimental framework, Zatorre et al. (Zatorre et al., 2007) report an experiment in which a delayed and distorted auditory feedback caused timing problems (for asynchronous feedback) and observable action selection. No timing change was observable when the pitch of the feedback was altered. This suggests that perception and action rely on one unique mental representation. The authors report that several brain regions are involved in timing and sequencing but notice the lack of study about spatial processing in musical tasks. Pfordresher and Dalla Bella (Pfordresher and Bella, 2011) showed that the effect of a delayed auditory feedback depends on both the introduced discrepancy itself and the motion dynamics (as well as the state of the movement in the sequence). Bravi et al. (Bravi et al., 2014) also observed that the acoustical features of an auditory stimulus can influence isochronous movement durations. Van Vugt (van Vugt, 2013) showed that expert musicians became quite independent from the auditory feedback referring to timing.

Figure 2.1 – A Dickcisse bird (*Spiza Americana*), from the family of Passeri, known as 'songbirds', in a demonstration of auditory-motor coupling.

The symptomatic lack of rhythmic synchronism (called 'beat deafness') is the inability for subjects to tap in synchrony on a musical beat[15]. Sowinsky and Dalla Bella (Sowiński and Dalla Bella, 2013) observed a 'mismatch' in the perception-action coupling in the case of subjects presenting abnormal rhythm synchronization and fair pitch detection in melodies. This observation suggests that the beat deaf symptom can come from a lack of auditory-motor mapping. Some testing procedure to evaluate rhythmical abilities have been developed and could be used to evaluate auditory-motor abilities of recruited subjects

---

[15] One has certainly experienced an awkward moment with his beat deaf neighbor during a concert.

(Farrugia et al., 2012).

Another ‘musical symptom’ has been studied which is the congenital disorder of impaired pitch perception, *amusia*. Subjects often have troubles to sing a targeted melody as well. This may be caused by the lack of perception, of course. But some exceptions can be found (Dalla Bella et al., 2009). In this study, the authors describe the case of two subjects suffering from amusia and being able to sign skilfully. They concluded supporting the idea of separate neural pathways for auditory perception and action. The auditory-motor loop of singing has been extensively studied using songbirds which show remarkable learning skills (Brainard and Doupe, 2000), especially about long-term learning, aging and social behavior (see a photo figure 2.1). In his review, Brainard underlines the similarities between songbirds and human speech or singing production. The curious reader will find a wide literature on that field.

Hickok et al. (Hickok et al., 2003) investigated brain activity when listening to speech and music, and using functional MRI were able to spot areas involved in auditory-motor development for speech production and musical abilities. An entire field of motor psychology is dedicated to speech learning and production, which we will not detail here.[16] (Zatorre et al., 2002; Houde and Jordan, 1998).

Let us mention here the case of focal dystonia among musicians, a particular neurological condition affecting muscles (typically hands or feet), and responsible for uncontrolled contractions and irregular postures and gestures. The remarkable work of Altenmüller (Altenmüller, 2003) and van Vugt (van Vugt, 2013) focused on understanding and treating musical dystonia, for instance with pianists and drummers.

Digital Musical Instruments (DMI)

The research field covering musical interfaces, Digital Musical Instruments and musical interactivity, has significantly grown since 2000. In particular, several new approaches “beyond the keyboard” and MIDI representations have been proposed (Miranda and Wanderley, 2006). The international conference NIME (New Interfaces for Musical Expression), started in 2001 as a workshop of the CHI conference (Bevilacqua et al., 2013a), contributed to expand an interdisciplinary community composed of scientists, technologists and artists. A competition of new musical instruments also exists since 2009 held at Georgia Tech[17].

As illustrated in Figure 2.2, a Digital Musical Instrument (DMI) can be formalized as composed of an *interface* or *gestural controller unit* and a *sound production unit* (Wanderley and Depalle, 2004; Miranda and Wanderley, 2006). These two components can be designed independently, in contrast to acoustic instruments. This representation must be completed by the *mapping* procedure which allows for linking the digital data stream, between the gesture features to the input of the sound processor, often represented as a data-flow chart.

---

[16] Music production and psychology fields remain very wide, and we cannot cover them entirely here.

[17] Margaret Guthman Musical Instrument Competition `http://guthman.gatech.edu/`

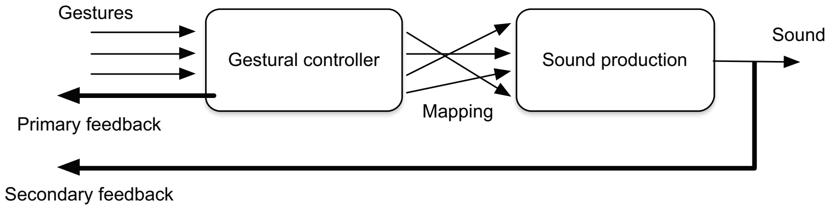


Figure 2.2 – Symbolic representation of a Digital Musical Interface. From (Wanderley and Depalle, 2004).

Examples of historical Digital Musical Instruments are presented in figure 2.3.

In (Wanderley and Orio, 2002), the authors postulate that input devices for musical expression respect two main trends : designing controllers that best fit existing motor control ability (e.g. imitating existing instruments) or designing so-called "alternate controllers" involving new gestural vocabulary (from a performance perspective). In all cases, different types of feedback mechanisms occur as illustrated in Figure 2.2: the primary feedback (visual, auditory and tactile-kinesthetic) and secondary feedback (targeted sound produced by the instrument). This way, feedback loops are created, and become central in the interaction.

In the field of musicology, acoustic and music technology, there has been an effort to formalize musical gestures (see (Godøy and Leman, 2010) for a review and in particular the review on sensorimotor control of sound-producing gestures (Gibet, 2009)). The field of neuroscience has also integrated an increasing number of studies on music and performance (see for example the four special issues of the Annals of the New York Academy of Sciences: The Neurosciences and Music (2003, 2006, 2009, 2012) that covers a large spectrum of this research). Nevertheless, to our knowledge, no study addresses directly sensorimotor learning in digital musical systems. This is mainly due to highly idiosyncratic use and design of DMI, and to the lack of repertoire and notation as found with acoustic instruments.

Most of the research work on DMI essentially addresses the design issues in building digital musical instruments, i.e. designing the gestural controller, mapping, sound production units and the interaction (Wanderley and Battier, 2000; Miranda and Wanderley, 2006; Paine, 2009; Paine, 2010). Research on musical gestures, including communicative gestures, sound-facilitating gestures or sound-accompanying gestures, has developed specific approaches. Several topologies have been proposed and point toward the fact that musical gestures cannot be reduced to simple discrete gestural control mechanisms, but involve both discrete and continuous movements, with different phases (*e.g.* preparation, stroke, release) and co-articulation effects (Rasamimanana and Bevilacqua, 2008).

One of the key problems in the DMI research field is the evaluation of the systems. Many approaches - including human-computer interaction inspired - were proposed but we deplore the lack concerning sensorimotor aspects. The notion of sensorimotor learning is

The Theremin, L. Theremin, 1918 *The Hands*, M. Waisvisz, 1984 *Biomuse*, A. Tanaka et al, 1992

Karlax, Dafact, 2004 *ReacTable*, S. Jorda et al., 2005 *Mo*, Ircam, 2010

Figure 2.3 – Examples of Digital Musical Instruments through the years. The *Theremin* uses the capacitance effect of the hands and two antennas to control the volume and pitch of an heterodyne oscillator. The *Hands* contain keys and movement sensors and control MIDI synthesizers. *Biomuse* is a controller based on biological signals such as EEG, EMG. The *Karlax* is a portable MIDI controller inspired from a wind instrument containing keys and inertial sensors and a Max/MSP mapping program. The *ReacTable* is an interactive table with objects representing distinct signal processes that can be moved to change the interaction between them, represented graphically. The *Mo* interface (Musical Modular Object) is a small wireless set of inertial sensors sending OSC messages allowing for gesture capture and object-related musical scenarios. The Mo, developed at Ircam, received the 1^{st} prize of the Margaret Guthman Musical Instrument Competition in 2011.

mainly ignored in that field while, from a behavioral point of view, this notion is central in the utilization of a gestural DMI because of the perception-action loop engaged. In the following paragraphs, we report works on the evaluation of DMI and the few works on auditory feedback that have been implemented in augmented instruments for educational purposes.

Evaluating digital music instruments Wanderley and Orio (Wanderley and Orio, 2002). noted that the design of input devices is generally associated to artistic projects with idiosyncratic choices, and generally lack of a general evaluation methodology. The authors proposed to use tools from the Human Computer Interaction field (HCI), and to define musical tasks that could be evaluated. Also inspired by the HCI field, Malloch et al. (Malloch et al., 2006) developed a design space for digital musical instruments. The design space is inspired by Rasmussen's theory of design (ecological interface design and SRK theory (Rasmussen, 1983)) and links Skill-, Rule-, Knowledge-based interactions to signal-, sign- and symbol-based interaction, respectively, for controlling music. Unfortunately, the authors do not propose evaluation methodology coherent with the proposed design space.

Human Computer Interaction methodology is not sufficient, since DMI are often evaluated based on their behavior during performances (O'Modhrain, 2011). Different qualitative methods of evaluation have been reported. In (O'Modhrain, 2011) the author proposed to integrate different types of evaluation according to the perspective from the performer, the audience, the designer and the manufacturer. Hsu et al. (Hsu and Sosnick, 2009) assessed the qualitative experience of musical systems by musicians and audience. Geiger et al. (Geiger et al., 2008) presented a study on participatory design of Theremin-like musical interface. They conducted an evaluation based on questionnaires for evaluating the hedonistic and pragmatic quality of interactive products. Poepel (Poepel, 2005) evaluates three digital string-based instruments together with a mapping thanks to a five point Likert-scale questionnaire by both professional and amateur musicians.

As already mentioned, quantitative evaluation of DMI, or of some elements constituting a DMI, requires the definition of musical tasks. Several main tasks have been defined and studied in the literature. Pitch selection and modulation have been evaluated with many different input devices like FSR (Wanderley and Battier, 2000), theremin, gloves, and Wiimote joystick (Geiger et al., 2008), but also accelerometer-based sonification systems like in (Fox and Carlile, 2005). In this study, the authors presented a system enabling to shape musical timbers with hand orientation, using OSC messages. They targeted a potential dance application at the time; now, many artistic applications of research on inertial movement sensors for dance can be found[18] (Bevilacqua et al., 2011). Navigating in acoustic features spaces is also a common task in musical control (*e.g.* timbers), as evaluated in (Vertegaal and Eaglestone, 1996) comparing mouse, joystick and a glove[19]. Targets are defined as coordinates in the chosen space, and evaluation measures the time to reach the target. Control integration though did not refer to a learning process. Triggering and rhythm production obviously depends on timing precision of the controllers, like in (Kiefer et al., 2008). Compared to electronic drum pads, inertial controllers proved to be harder to use because of the lack of sensory feedback in the gesture. Timing accuracy of drum pads has been benchmarked in (Collicutt et al., 2009). Multi-modality, integrating haptic feedback from drums learning, has been proposed in (Holland et al., 2010). The authors showed that beginners are able to learn intricate patterns from haptic stimuli alone.

The mapping procedures have been formalized and recognized as key elements in the digital instrument design, with both technical and artistic problematics (Hunt et al., 2003). It has been measured that users prefer more complex mappings once they have been learned because they allow for more control and expressiveness (Hunt et al., 2000; Hunt and Kirk, 2000). Unfortunately, evaluation is often based only on user experience, through questionnaires (Merrill and Paradiso, 2005) or directly on stage (Collins, 2011).

---

[18] Greg Beller - Synekine Project - `http://www.gregbeller.com`; Richard Siegal - `http://www.thebakery.org/`; Emio Greco, Double Skin/Double Mind Project `http://www.researchcatalogue.net/view/?weave=1585`. See also video archives of dance and electronics pieces on the Ircam catalog at `http://ressources.ircam.fr`.

[19] `http://thepowerofglove.com/`

Auditory feedback and sonification for instrument practice A small number of works have been conducted on interactive systems providing auditory feedback during instrument practice, based on either sound analysis, MIDI data or movement analysis. Ferguson (Ferguson, 2006) reported different interactive sonification algorithms based on the real-time analysis of the sound (note onset, rhythm, loudness control, legato, and vibrato) to provide KR auditory feedback to singers and instrumentalists. In the context of the iMeastro project (EU-IST), Larkin et al (Larkin et al., 2008) and Rasamimanana et al (Rasamimanana et al., 2008) implemented several approaches of auditory feedback for string players based on the bow motion. Grosshauser and Herman (Grosshauser and Hermann, 2009) also proposed different systems to provide multimodal feedback to violin players. Hadjakos developed sonification feedback for piano players and reviewed existing systems providing visual or auditory feedback for pedagogical applications. Promising results were obtained but no large-scale evaluation have been conducted to assess these systems about sensorimotor learning.

2.3.3 Auditory feedback for sport and physical exercise

Sport training offers a particular case study for advanced motor control over a wide variety of actions. One can observe complex movements performed with a high level of performance and precision among experts. Sport context implies situations with body control alone (like gymnastics (Baudry et al., 2006)), or with accessories (pole vault), one opponent (table tennis) or collective interactions (team sports like rugby (Correia et al., 2011)), providing a rich experimental framework. Significant efforts are made to improve performance in this field, leading most of the time to training procedures and more rarely to fundamental research results. However augmented feedback is quite investigated as athletes are already used to analyze quantitatively their performance on training machines. Furthermore, tasks are often repetitive and subjects usually show robust and low-variance repeatability[20]. Athletes show indeed reduced variability on key movement features (see the example of table tennis stroke in appendix A). These specificities can be detrimental for fundamental research, because trained athletes outperform most of the sensorimotor tasks.

In this section, we report examples of studies which considered sonification to improve particular movement features in a sport context (there is a substantial number of dedicated journals on sport studies). A pilot study presented in appendix A, describing an instrumented table tennis racket for sonification purposes, showed two issues when considering movement sonification for specific sport gestures. Technically, and because of practical constraints, it can be very difficult to sense the desire feature and to deliver a sound feedback which is informative enough. Second, we wonder whether one can defined an ideal gesture, which could be used as a reference to generate a KR type feedback. Some studies we report faced these issues too. Nevertheless, we cite experimental results that clearly

---

[20] In musical acoustics, musical robots are sometimes used to ensure normalized and robust repeatability while playing an instrument.

emphasize the benefits (and great potential) of using sound as an augmented feedback for sport training.

Karageorghis and Terry (Karageorghis and Terry, 1997; Terry and Karageorghis, 2006) underlined the benefits of music in healthy patients when practicing sport or activities. A better synchronization of rhythmic activities and a better feeling about physical exercise has been noticed. Hoffmann et al. suggested that running in synchrony with music could reduce the variability of the stride (Hoffmann et al., 2013) and increase the locomotor-respiratory coupling (Hoffmann et al., 2012). They also found that the energy consumption is reduced when the rhythmic feedback matches the preferred stride frequency of the subjects, suggesting a need to adapt the system to inter-individual variability.
As music can act as a perceptual diversion from painful or tiring proprioceptive sensations during physical exercise, Fritz et al. (Fritz et al., 2013) observed that musical agency (Knoblich and Repp, 2009) could reduce the perceive exertion during physical workout. Takeuchi observed that depriving tennis players from audition led to decreased performance in the game, especially receiving services. This confirms the close link between auditory perception and action in sport - where extra auditory information from the scene or the opponent play is extremely valuable.

It has been observed that inter-limb coordination while juggling can be improved by movement-driven audio feedback (Bovermann et al., 2007; Zelic et al., 2011), exploiting the advantages of audio on sight and multimodal feedback. Wellner (Wellner and Zitzewitz, 2008) also found that the short processing time of auditory information could help participants in tasks where synchronization and speed are required (obstacle overcoming while walking on a treadmill with virtual reality). Sound was judged more helpful than visual cues in that case and improved gait speed. The help of auditory feedback has also been reported in gymnastics (Baudry et al., 2006). In this last case, a two-day retention test was performed and revealed no loss in performance of subjects trained with auditory feedback.

A specific task has been studied in (Effenberg et al., 2011; Wolf et al., 2011; Sigrist et al., 2014; Schaffert et al., 2011), either with virtual reality or in the real environment. In these cases, the authors chose the rowing task which requires control over a large number of degrees of freedom, kinematic and dynamic control as well as physical strength, and manual grip. In addition, it implies cyclical and repetitive movements, which are easier to sonify and perceive. As a result, it stands as a rich experimental context. The sonification of oar kinematics and dynamics proved to be valuable for movement perception and error reduction in these studies. Both the sonification of errors (the distances to a reference pattern) and of kinematic variables proved to be efficient. The authors emphasized the need for more research on sonification of physical activities (which are also very dependent on the context) or daily life activities[21].
An example in speed skating has been presented by Stienstra et al. (Stienstra et al., 2011).

---

[21] Neurophysiological studies are also needed in this field.

As the rowing task, this activity is based on regular and repetitive movements, which can easily benefit from sonification. This study has been carried out with a design approach - only one subject has tested the device. They used a filtered pink noise to sonify the pressure applied on front and rear supports of each skate blade. The skater showed more regularity with the auditory feedback, although learning has not been properly assessed in this study. Virtual reality has also been used in (Todorov et al., 1997) to evaluate motor learning of complex table tennis moves. Subject had to learn complex movements in this virtual environment, and performance was then tested in a real task. Subject who trained with auditory feedback in this environment performed better than those trained in a real-task practice (with a comparable amount of training according to the authors). The potential transfer between virtual and real task was also evaluated. The authors noted the individual variability between the subjects when they tried to imitate the coach movements. In the case of table tennis, they hypothesized that learning a complex motor task could be done in two steps, a- learning general requirements of successful movements and b- learning the easiest movement satisfying these requirements. Besides, virtual reality enabled to superimpose the coach's racket movement to the one of the subjects, which might be useful for training complex motor movements.

2.3.4 Auditory feedback for rehabilitation

Stroke is one of the most important cause of impairment today. It often leads to typical physical pathologies like hemiparesis (motor and sensory impairment or paralysis of limbs), apraxia (loss of know-how with objects), aphasia (inability to talk) or hemianopsia (loss of half of the visual field) (Maulucci and Eckhouse, 2001; Robertson et al., 2009). Besides, it is a real public health issue worldwide (Loureiro et al., 2003). The goal of motor rehabilitation is to restore the ability to be independent regarding activities of daily living (ADLs) after a stroke. This is essential for the patients to re-integrate into social and domestic life (Avanzini et al., 2009). In this section, we will report on studies and methods investigating auditory feedback in rehabilitation therapy processes[22]. This application will also be discussed as an important perspective in section 8.5.

After the acute phase of the stroke (from $t0$ to $t+3$ months) where intensive care are lavished, recovery continues in the sub-acute ($t+3$ to $t+6$ months) and chronic phases (after $t+6$ months). It is noticeable that major recovery of upper limb extremity occurs in the early months (Maulucci and Eckhouse, 2001), and that many studies advise to start rehabilitation soon and intensively after stroke. However Maulucci and Eckhouse also suggested that a specific rehabilitation on a motor function should be done after 6 months once patients have reached a plateau in their recovery, to ensure improvements are not due to early spontaneous recovery. Van Vliet and Wulf (van Vliet and Wulf, 2006) suggested indeed to adapt the interactions parameters (*i.e.* feedback) depending on the

---

[22] We claim no comprehensiveness in this overview, as the literature is very specific and quite undisclosed.

stage of rehabilitation of the patients. Cirstea et al. (Cirstea et al., 2006) added that stroke severity and possible cognitive impairments must be taken into account when setting up rehabilitation process.

Augmented feedback for rehabilitation

During rehabilitation, physiotherapists always use augmented feedback (Hartveld and Hegarty, 1996). Three main ways to provide it during rehabilitation are specified in this review: personally (verbal or non-verbal), through equipment to enhance intrinsic feedback (accessories, tools) and equipment to give augmented feedback (mainly electronic devices). Feedback can sometimes also be related to the benefits of the task or a form of reward. The three modalities of feedback were compared in (Hartveld and Hegarty, 1996): the first one proved weak and slow, the second improved the functional realism of the task, and the last proved accurate, immediate and quantitative. The authors concluded by giving necessary conditions for efficient augmented feedback during rehabilitation: it has to be “rewarding, desirable, plentiful”, related to kinematics and kinetics of the movements, functionally related and precise.

Using extrinsic feedback to help rehabilitation is mainly motivated by the great plasticity of the nervous system that can be beneficial to stroke survivors: “Adaptive plasticity linked to rehabilitation is predicated on the hypothesis that short duration connections achieved through fast Hebbian learning facilitate the establishment of more durable connections occurring with repeated practice” (Subramanian et al., 2010) (see also section 2.1).

Subramanian et al. presented a review concerning post-stroke upper limb motor learning aided with extrinsic feedback (Subramanian et al., 2010). Although only one study they detailed used auditory feedback, this systematic review concludes that evidence can be found that extrinsic feedback is beneficial for motor learning in post-stroke rehabilitation. They explain that, in rehabilitation, knowledge of performance (KP) is used predominantly. A frequent provision of knowledge of results (KR) can however improve performance but can disturb long term memorization of a task (see also (Engardt, 1994)).

Implication of the subject and motivation are important to patients in order to focus on the rehabilitation exercise. Avanzini (Avanzini et al., 2011) states that finding the strategies to enhance engagement and motivation in motor rehabilitation is an open research challenge.

Motivations

In (Karageorghis and Terry, 1997) the authors explain that exercises are facilitated by synchronization with musical accompaniment. Music can also reduce the feeling of effort during exercise (Fritz et al., 2013). Finally, the great variety of styles and atmosphere offered by music can, if chosen well, enhance enjoyment levels and adherence to physical activity. However it is still unknown how this mechanism can work in motor rehabilitation (i.e. with impaired and often shell-shocked patients). The multiple effects of music-based therapy on learning and engagement are also underlined (Thaut, 2005). Thaut stated that the available background on neurologic music therapy (including rhythm) has established an entire discipline of music therapy practice.

It is generally recognized that feedback should have a multimodal form to be more effective (Avanzini et al., 2011). Multimodal feedback can improve the performance in a complex motor task but the potential of auditory feedback is still underestimated. Avanzini et al. proposed that auditory feedback for rehabilitation in ADLs should be used along with other modalities, and particularly to sonify the user's movements and the environment. One of the strongest advantages of auditory feedback is that it is easily perceivable without requiring much attention or strong cognitive effort (Avanzini et al., 2011). Furthermore, for patients who lay in bed for a long period of time and suffer from a lack of attention, auditory feedback is a practical way of augmented feedback.

Many questions still remain unanswered and some authors, like van Vliet and Wulf (van Vliet and Wulf, 2006), deplore that most of the existing research on the subject only evaluate the effectiveness of devices and do not try to answer more fundamental (or theoretical) questions about KP vs KR effects or feedback integration mechanisms. Few studies observed temporal parameters of feedback in stroke patients (frequency, summary, average, delay,...) and none with auditory feedback. Observing and understanding the mechanisms involved in auditory-motor learning on healthy subjects are essential milestones towards application to patients. Nevertheless, stroke patients may suffer from sensory lesions or lack of concentration compared to healthy subjects. Thus, it is not clear whether patients recovering from stroke can learn with the same processes as healthy subjects. Their intrinsic feedback paths may be damaged, giving even more importance to providing extrinsic feedback. As the processing of implicit versus explicit feedback can be changed, and difficulty of tasks will differ from healthy patients, changing feedback parameters should have different effects on patients (van Vliet and Wulf, 2006). The transposition of models and experiments from healthy to stroke patients will have to be carried out carefully, if possible.

A word on biofeedback

Biofeedback can be defined here as the use of instrumented devices to provide physiological information about the movement. When the information channel is audition, biofeedback means sonification of body physiological data, using EEG, EMG, body kinetics, ECG, fMRI, etc. (Baier et al., 2007; Pauletto and Hunt, 2009) (see 2.2). Originally, audio biofeedback was studied along with visual feedback and used very simple alarm-type sounds (Dozza et al., 2007). Biofeedback has been mainly used for gait and posture tests or in children with cerebral palsy, but the audio modality is still under-studied.

Upper limb movement

Hemiparesis (impairment or paralysis of the right or left side of the body) is a major cause of disability after a stroke. Upper limb dysfunctions impair most of the ADLs and cause a loss of independence therefore can lead to social exclusion. Upper limb rehabilitation appears then as a major issue in rehabilitation after a stroke.

A frequent type of test concerns reaching gesture with the arm. Most of the rehabilitation processes include this procedure which is fundamental for recovery of ADLs. One of the first studies that used auditory feedback in rehabilitation of upper limb was published by Eckhouse (Eckhouse et al., 1990) in 1990. Hemiparetic subjects received KR in the form of tones indicating subject's score and precision in a reaching task of a target presented on a touchscreen. The test group had better performances than the control group due to auditory feedback. Eckhouse concluded that enhancing recovery of stroke patients with specific and regulated feedback on guided limb motion is proved. He adds that "the realization of significant modification demands utilization of specific feedback presented to sensory systems capable of participating in the restorative process", underlining the importance of the design of the feedback and understanding of the mechanisms involved.

The task of reaching with hemiparetic patients has been tested with real-time sonification in (Maulucci and Eckhouse, 2001). The task consisted in touching targets situated in front of the subjects. They were given the deviation from the normal path they were following with their hand (acquired with a magnetic hand-tracking device) and the magnitude of this error through frequency and amplitude modulated sounds. This spatial sonification was found to help to "visualize" the correct path by the subject. The authors also noticed that some subjects were lured by the final target and less focused on the path to follow; this raises the question of attention with augmented feedback and how to design correctly the appropriate sound feedback. In order to focus training on ADLs, the authors proposed that an other task should be added at the end of the gestural path (such as pressing a button); the whole reaching strategy may thus be modified. In addition, they detailed many gesture kinetics and kinematics parameters they recorded to evaluate movement and learning in such a reaching test. Robertson and colleagues (Robertson et al., 2009) carried out a study on reaching movements with auditory feedback on hemiparesis patients, comparing the brain hemisphere affected and two types of auditory feedback (volume and stereo panning). Their main conclusions stated that the effectiveness of the auditory feedback depends on the hemisphere affected by the stroke. Patients with the left hemisphere damage even showed worse results with the feedback. One hypothesis is that lesions in the left hemisphere may have disrupt feedback processing capacity. The authors also suggest that auditory feedback would be more appropriate for giving temporal information and KR rather than KP (better for visual feedback).

Music has been one of the feedback form explored[23]. Ghez and colleagues (Ghez et al.,

---

[23] We will not focus on musical therapy *per se* here.

2000) explored musical scales for movement sonification to improve inter limb coordination in patients suffering from proprioception deficits. They showed encouraging results using sonification of joint motion and timing cues, but without control experiment.

Studies previously mentioned in section 2.2 sometimes warn about the cognitive overload that additional auditory feedback may cause during a complex motor task. In the case of the control of a prosthetic hand, users have to rely extensively on visual input. Gonzales et al. (Gonzalez et al., 2012) showed that auditory feedback along with vision on prosthetic's kinematic and grasp actions reduced the attentional demand when performing daily activities tasks. The results were obtained with EEG, EMG, breathing, electro-dermal measurements and questionnaires.

In a recent paper (published in October 2014) by Scholz et al. (Scholz et al., 2014), the lack of established sonification-based rehabilitation strategy is still deplored. The authors presented a study where they sonified the 2D mouse pointing movements of elderly subjects, using pitch and brightness. Subjects were first familiarized with the auditory mapping in a free exploration phase, and then heard sounds corresponding to specific coordinates on the screen (in the pitch/brightness 2D space). The task was then to point the coordinates that would produce the same sound. Results showed that pitch variations led to better precision and that learning was faster if pitch was mapped on the vertical direction. Another recent article presented a movement sonification system based on inertial sensors for rehabilitation (Schmitz et al., 2014). They proposed a reaching and grasping task to seven post-stroke patients. Performance was assessed with standard rehabilitation tests (box and block test and nine-hole peg test). No significant results were obtained (due to large inter-individual differences), but tendencies are again in favor of positive effect of movement sonification.

Gait, posture and rhythm

Auditory and vestibular information are transmitted to the brain through the same nerve (Dozza et al., 2007). It is likely that postural alignment is subconsciously influenced by auditory information: this stands as a rationale for using auditory feedback in balance rehabilitation, according to Dozza et al. Another motivation can be found in (Easton et al., 1998) where the authors showed that sound delivered with two lateral speakers can reduce center-of-pressure sway in congenitally blind and sighted people, indicating that auditory perception can be easily integrated for gait control. In (Dozza et al., 2007) the authors tested the influence of an audio biofeedback system on postural sway in healthy and vestibular loss patients. A continuous low-volume sound was emitted and modulated when the subjects were swaying outside their natural range (previously measured). The sounds used were pitch and amplitude-modulated stereo sine waves. Procedure evaluated the sway eyes closed, then eyes opened with foam under their feet or not. The conclusion is that the subjects used the extrinsic feedback all the more that their sensory information was reduced. A variability of weight given to the different sensory information was also noticed.

Baram and colleagues (Baram and Miller, 2007) tried to evaluate the residual short-term effect of auditory feedback during gait rehabilitation. Fourteen multiple-sclerosis patients with gait disturbance (mainly due to cerebellar ataxia) participated. Their results suggest that a positive and rewarding auditory feedback on steady balance gait can improve walking parameters, all the more that the patient's baseline speed was low. The feedback may sometimes have negative effect on healthy patients like disturbance. Nevertheless, the authors underlined that the inner mechanism between auditory signals and coordinating movements remained to be found and that studies with larger groups were required.

In (van Vliet and Wulf, 2006) the authors noted that auditory feedback improved performance in sit-to-stand movements after a week of training (greater symmetry in body-weight distribution). An experimental study by Batavia et al. corroborates this point (Batavia et al., 1997). A pressure sensor in a stroke patient's wheelchair cushion triggered a buzzer sound when the weight was not properly distributed. The goal was to make the patient sit correctly on his wheelchair and stay straight. The authors reported an improvement after two to three days in the patient's symmetry in weight and midline perception. After seven weeks the patient was able to sit alone and improved some dynamical controls of his trunk, head and shoulders. Although this case concerned only an isolated 74-year-old patient, long term adaptation effects on daily movements can be observed with auditory feedback.

Engardt (Engardt, 1994) carried out a study to assess long term effects of auditory feedback on sit-to-stand body weight distribution in patients with a paretic leg. It is noticeable that it is the only study we found which studied long term retention effects of auditory feedback in rehabilitation. Patients (feedback and control groups) were trained during six weeks to sit-to-stand movements. Thirty-three months later, the results showed that patients in the feedback group had lost their relearned tasks performance after this time more than the control group did. However they were faster to accomplish them. Engardt proposed several hypotheses such as patients mostly used their healthy leg for speed and security after the tests. The author concluded by stating that an auditory feedback should be delivered with reduced frequency, in very long sessions (maybe thousand times according to Bach-y-Rita and Baillet (Bach-y Rita and Baillet, 1987)), over long periods of time and during open task situation - as the closed situations seemed not similar enough to a real environment for the patients.

A study tested auditory feedback on sway, with isolated otolith disorders patients - who have been found to respond weakly to regular vestibular rehabilitation strategies (Basta et al., 2008). Thirteen subjects received audio from the Sway-StarTM system that produces a single tone in 3 exercises: standing eyes closed and standing on foam eyes opened then eyes closed. The system is attached to the patients lower trunk and senses trunk angle velocity in roll and pitch planes. A tone was emitted when the angles crossed a threshold; the tone was emitted from one (out of the four) loudspeaker towards the patients who had to correct their sway. The exercises were performed everyday during two weeks. 85% of patients on the test group showed significant decrease of trunk sway, most significantly when walking on the foam.

Inertial measurement units with accelerometers have been also used to sonify gait movements (Chiari et al., 2005). In (Batavia et al., 2001) the authors tested on a single patient (12.5-year-old) a membrane switch buzzer in the shoe to study gait. The authors underlined the ability of the patient to walk on her own after 3.5 weeks of training with auditory feedback but this study focuses on a particular case.

The temporal and rhythmic aspect of musical patterns has been studied as a good candidate to help synchronize movement and regain lost regularity. Auditory feedback allows for designing salient temporal features, that are thought to facilitate gait and cadence rehabilitation (Thaut et al., 1996; Thaut et al., 1997). Whitall et al. (Whitall et al., 2000) gave the two main advantages of a bilateral rhythmic auditory cueing system (BATRAC). First, evidence suggest that motor learning could be enhanced with bilateral movements (simultaneous or alternating): both arms are linked in one control unit in the brain and experiments showed that learning a task with one arm can lead to a transfer of skill to the other arm. Second, this protocol uses a rhythmic repetition that is a classic learning principle. A constant frequency leads the motor system to acquire a certain regularity (Thaut et al., 1996). Also, synchronizing the end of a gesture with a discrete sound gives an attentional goal to the patient during the exercise.

A musical game has been developed by van Wijck et al (van Wijck et al., 2012), using personalization music choices and Wiimote controllers as displacement sensors. They proposed a tap/tempo game with a graphical interface that indicates beats and bars of the music that patients had to point with the Wiimote. Audio gaming rehabilitation systems can afford qualitative and quantitative evaluation, as well as therapeutic inputs and outcomes. This technique, though, relies importantly on playfulness and enjoyments of the design.

Parkinson's patients are receiving a growing interest to develop new training protocols. One of the important path explored is the use of rhythmic auditory feedback or stimulation to help gait and stride. Mcintosh (Mcintosh et al., 1997) used a rhythmic auditory stimulation (RAS) on Parkinson's patients, first by matching their own baseline cadence and then increasing the rate of the cues (10%). Retention was also later tested without the device. The results suggested that velocity, stride length and cadence can be improved with increasing-rate RAS. Recent studies focused on the ecological approach of the action-sound relationship in order to design promising protocols (Young et al., 2014). Serious gaming for Parkinson's patients is also a novel approach showing increasing attention (so is virtual reality, see below for examples) (Paraskevopoulos et al., 2014).

These results tend to show the potential of auditory feedback on gait and posture rehabilitation, whether it is delivered from weight distribution or external cues and rhythmic stimulation. These results, even though quite sparse in the literature, show the positive effects of sound feedback for rehabilitation through biofeedback systems. That being said, there are still very few results available on long term follow up and assessing inter-individual variability. The lack of theoretical and experimental frameworks demand much further work in this domain.

Robot-assisted rehabilitation with auditory feedback

A sizeable part of studies carried out with auditory feedback for rehabilitation concerned robot-assisted rehabilitation devices. Technological developments allowed for the rehabilitation and neuroscience fields to benefit from robotics, which became quite recently an important branch of technology-assisted rehabilitation. The implementation of robot-assisted rehabilitation is still ongoing but remains expensive and unwieldy. The main advantage of robotic systems is that motion can be easily captured and augmented feedback is pretty easy to integrate as servomechanisms are already involved. Unfortunately, the actual benefits of the feedback on motor-learning are also poorly assessed in this case; only few studies integrated sound to robotic devices in that purpose.

In a review, Avanzini et al. took an inventory of 47 papers describing 36 robot-assisted rehabilitation systems (Avanzini et al., 2009). The authors focused on whether the auditory modality was used and what type of sound: earcons, auditory icons, speech or sonification. The first conclusion is that a majority of system does not use audio at all. Most of the time, sound takes the form of auditory icons and earcons and the authors underlined that only a few systems use sonification, showing that the potential of interactive and continuous auditory feedback is still underestimated. See also (Sigrist et al., 2013) and references cited in (Avanzini et al., 2009) for further reading.

Sound is often used as a rewarding KR and to improve the user's engagement in technology-assisted rehabilitation systems, as in (Cameirao et al., 2007). The patient is rewarded with a "positive sound" whenever he succeeds in a specific game. Discrete auditory feedback in stroke rehabilitation using robotic manipulation is also presented by Colombo (Colombo et al., 2005) where the device provided visual and auditory feedback to the patient to signify the start, the resting phase, and the end conditions of the exercise (no details are given on the sounds). Other work in rehabilitation using robot assistant was done by Secoli et al. (Secoli et al., 2011). The authors showed that a simple sound feedback (beeps with frequency increasing with the tracking error) enabled the participants to perform a tracking task in the presence of a visual distractor. Task accuracy was computed as the difference in position between sessions with and without the distractor. The potential of real-time auditory feedback on performance errors in robot-assisted rehabilitation systems in emphasized in this study. Similar results are found in (Rosati et al., 2011).

In their review, Avanzini et al. (Avanzini et al., 2011) deplored that few cases of technology-aided rehabilitation have been transferred to real-world application in a medical context, which is an important objective of rehabilitation research today.

Virtual environments for motor rehabilitation

A derivative of robot-assisted systems is the use of virtual reality environments - also allowed by the development of fast real-time interactive systems. For reviews on rehabilitation in virtual reality context, see articles by Sveistrup (Sveistrup, 2004) and Holden (Holden, 2005). Multimodal interaction and virtual reality have been explored, for instance in (Loureiro et al., 2003); the complexity of the setup proposed makes the potential transfer to home-based rehabilitation, or even small clinics, quite difficult. Only a few articles referenced deal with sound or auditory cues (Shing et al., 2003). Most of the time, audio is used to enhance spatial dimensions, such as orientation and localization (Holden, 2005). Multiple loudspeaker systems are often used to create "spatial" rendering and sensory immersive environment. Audio processing in virtual reality systems answers efficiently to a need for realism of these systems.

Lehrer et al. (Lehrer et al., 2011a; Lehrer et al., 2011b) wrote a large interesting article to conceptualize multimodal interactive upper limb rehabilitation and to give systematic guidelines to design such systems. They focused on adaptive and interactive rehabilitation by the mean of phenomenological approaches and embodied knowledge brought into rehabilitation methods. The system they presented is based on motion capture and can be adapted to particular rehabilitation needs through multimodal feedback (sound and video animations). Gaming and game metaphors are also of interest for virtual rehabilitation environments (Correa et al., 2007). Multisensory platforms and gaming allow for motivation and engagement, which is critical for the success of a rehabilitation process. Cameirao et al. (Cameirao et al., 2007) set up a direct interaction virtual reality system where audio had a rewarding function. Stroke patients with left hemiparesis were tested using a tracking camera in a gaming environment, called Rehabilitation Game System RGS. Each time the patient (with or without paresis) accomplished the goal of a specific game, a "positive sound" was triggered. Feedback was only a rewarding KR. The system is based on the hypothesis that "motor execution combined with visual feedback can trigger the mirror neuron system and thus give access to the central motor system". Continuous auditory feedback can also be used for realism but it is not directly related to movement, and rather to the object which is being manipulated (Johnson et al., 2003; Boian et al., 2003; Nef et al., 2007).

2.4 General comments

By browsing through the available literature on the use of auditory feedback for movement control, several issues can be noticed. The literature is scattered over various fields (neuroscience, psychology of perception, sports sciences, music technology, rehabilitation, etc.), and is quite patchy. The use of sonification appears more as a *technique* than a *discipline* in itself. The lack of result-based consensus and methods prevents sonification from being more generalized. For instance, some results found are contradictory (*e.g.* on time exposure of KR in a learning task or about the efficiency of error-related feedback).

Studies are often based on inside-laboratory prototypes, dependent on the context and the tasks that are proposed. The evaluation of the systems is not globally formalized, often delicate and sensitive to inter-individual variability (somewhat like in digital musical instrument research). Very few studies focused on long-term learning (more than a few days). Many papers conclude - as often - with the statement that further studies are necessary to deeper investigate the auditory-motor coupling in the various fields. Most of the studies barely use the great potential of auditory feedback they tend to underline. The practical applications outside the laboratories - which are actually being imagined and look exciting - demand further research, including theoretical basis.

The gesture/sound mappings used are often very simple and direct, but still could prove effective. Some general results on specific characteristics of auditory feedback and sonification are available, and found coherent between different studies. For instance, the *temporal precision* and reactivity of the auditory feedback has been shown to be of critical importance. The *dynamic* nature of the feedback seems to make it more usable and to ‘tighten’ the action-perception loop. If a cinematic variable is sonified, the perceived features are often the *variations* of this variable. Several studies underlined the importance of the spectral content of the auditory feedback, which enrich the perceived features and seem suitable for movement sonification; this is probably due to the important ecological perception humans have developed of spectral dynamics, from sounds produced by physical interactions or mechanisms. Regarding the tasks studied, it appears that the difficulty proposed should be important to ensure room for improvement so that the effects or the feedback are maximized.

It can be noted that there is, as we speak, a growing interest in the use of auditory feedback for motor learning. The number of related papers on the subject is visibly increasing and in diverse scientific communities. Rehabilitation in the case of stroke or neurodegenerative diseases like Parkinson’s is particularly targeted today, as it can be seen in perception and psychology studies (Young et al., 2014). This recent momentum is encouraging and augurs important applications of movement sonification.

In the light of these (still to confirm) results, several paths to follow, as well as multiple important aspects emerged (context, task, mapping etc.), rather than one precise question. We then chose to explore different experimental contexts, corresponding to the next chapters. As a result, our questions also included methodological and task design points. The next chapters thus investigate the use of dynamic and continuous sonification of motion in different tasks, that will bring complementary observations and results. Importance was given to ecological relationship between movement and sound, as well as both practical and fundamental issues regarding future applications of movement sonification.

Chapter 3

From ear to hand: the role of the auditory-motor loop in pointing to an auditory source

E. O. Boyer[1,3], B. M. Babayan[1], F. Bevilacqua[1], M. Noisternig[1], O. Warusfel[1], A. Roby-Brami[2], S. Hanneton[3], I. Viaud-Delmon[1]
[1] IRCAM, STMS-CNRS-UPMC, Paris, France
[2] ISIR, UPMC CNRS UMR 7222, Paris, France
[3] LPP, Paris Descartes CNRS UMR 8242, Paris, France

Refereed article published in *Frontiers in Computational Neuroscience*, 7(April):26, 2013.

Abstract: Studies of the nature of the neural mechanisms involved in goal-directed movements tend to concentrate on the role of vision. We present here an attempt to address the mechanisms whereby an auditory input is transformed into a motor command. The spatial and temporal organization of hand movements were studied in normal human subjects as they pointed towards unseen auditory targets located in a horizontal plane in front of them. Positions and movements of the hand were measured by a six infrared camera tracking system. In one condition, we assessed the role of auditory information about target position in correcting the trajectory of the hand. To accomplish this, the duration of the target presentation was varied. In another condition, subjects received continuous auditory feedback of their hand movement while pointing to the auditory targets. Online auditory control of the direction of pointing movements was assessed by evaluating how subjects reacted to shifts in heard hand position. Localization errors were exacerbated by short duration of target presentation but not modified by auditory feedback of hand position. Long duration of target presentation gave rise to a higher level of accuracy and was accompanied by early automatic head orienting movements consistently related to target direction. These results highlight the efficiency of auditory feedback processing in online motor control and suggest that the auditory system takes advantages of dynamic changes of the acoustic cues due to changes in head orientation in order to process online motor control. How to design an informative acoustic feedback needs to be carefully studied to demonstrate that auditory feedback of the hand could assist the monitoring of movements directed at objects in auditory space.

Keywords: Spatial Audition, Human, Pointing Movement Kinematics, Orienting Movements, Reaching, Auditory-motor Mapping, Movement Sonification.

3.1 Introduction

Interactions between the auditory and motor systems are mainly studied in the context of musical rhythm or vocal sounds perception and production (*e.g.* (Hickok et al., 2003; Chen et al., 2009)). However, hand pointing to sounds is often used to study auditory localization. It is a complex task that relies on a precise representation of auditory space that can be used for the control of directional motor output. Just like pointing to visual targets, it involves different modular neural processes since spatial information about the target position and hand position have to be combined across different senses and reference frames.

In order to address the mechanisms whereby an auditory input is transformed into a motor command, we studied online auditory control of the direction of pointing movements towards auditory sources. We first investigated whether pointing movements were more accurate when the target was present throughout the entire pointing movement than when the target disappeared shortly after the hand movement had begun.

We then added an auditory feedback of the pointing hand's position during the entire hand movement to evaluate whether human subjects could use such a feedback. This additional auditory feedback named auditory avatar (by analogy with avatars used to represent visually a part of the body of a participant in a virtual environment) was used in order to evaluate whether it would constitute stable and relevant information to guide the motor action of the user, as already suggested by recent results indicating that auditory information is used to control motor adaptation (Oscari et al., 2012). With such an auditory feedback, the auditory modality conveys supplementary sensory information that is correlated with proprioception and set in modular processes in the same spatio-temporal reference frame as the target, hence facilitating precision in the pointing task. A well-designed auditory avatar, which corresponds to a sonification transforming relevant parameters of human movement patterns into appropriate sound, could be used to enhance perception accuracy and would be useful for sensory substitution and motor training technologies.

The first auditory avatar condition was contrasted to a shifted condition where the heard hand position did not correspond to the actual hand position thus resulting in a discrepancy between auditory and proprioceptive information. Similar methodology can be found in (Forma et al., 2011), where participants were asked to point to virtual targets in a spatialized audio environment using the openAL library (interaural time and level differences based audio environment). Studying online adaptation to this sensory conflict was expected to provide further information about the contribution of auditory inputs generated by arm movements to motor control.

3.2 Materials and methods

3.2.1 Subjects

Twenty-four self-reported right-handed volunteers (12 females and 12 males; 25.6 ± 6.6 years old) participated in the experiment. All were healthy and had normal hearing. The study was carried out in accordance with the Declaration of Helsinki. All subjects gave written informed consent and were paid for their time.

3.2.2 Experimental setup

The experiment used real-time controlled virtual audio rendering for both representing sound sources at the target positions in space and attaching sounds to the subject's right hand during the pointing movement. Audio was played back over headphones and subjects were seated in front of a table from which the auditory targets virtually originated. To prevent any visual input interference during the experiment all subjects were blindfolded.

The stimuli for target sources and the auditory avatar were (mutually uncorrelated) white Gaussian noise signals. The virtual audio targets as well as the auditory feedback of the hand position were provided with the Head-Related Transfer Functions (HRTFs) binaural technique (Wightman and Kistler, 1989a; Wightman and Kistler, 1989b)). Spat∼, IRCAM's software for real-time sound source spatialization, was used to create the binaural signals. Binaural rendering uses HRTFs to reproduce the sound pressure at the ear entrance that corresponds to a sound source at a given position in three-dimensional space. Processing a monophonic audio signal with a set of HRTF filters and playing these signals back over headphones creates the illusion of a virtual sound source at the corresponding position in space. The spatialization of the sounds (stimuli and hand position) was calculated in real-time through the tracking of the head's and right hand's positions and orientations using a six-camera Optitrack (by Natural Point) 3-D infrared motion capture system. To this end, two rigid sets of markers were placed on the headphones and the right-hand's forefinger. They were respectively composed of seven and four reflective markers tracked by the cameras. The coordinates of the hand and head's locations in space were measured and recorded with the tracking system at a sampling frequency of 100 Hz. The minimal latency of the overall system is then 10 ms - with an audio latency of 0.6 ms - which is fast enough to ensure perceptive coherence when localizing virtual sound sources (Brungart et al., 2004). The orientation of the 7-marker rigid body fixed to the headphones allowed for computing the heading direction (0° is forward, positive is to the right, see figure 3.1. The endpoint used to measure the kinematics of the hand corresponded to the tip of the index finger.

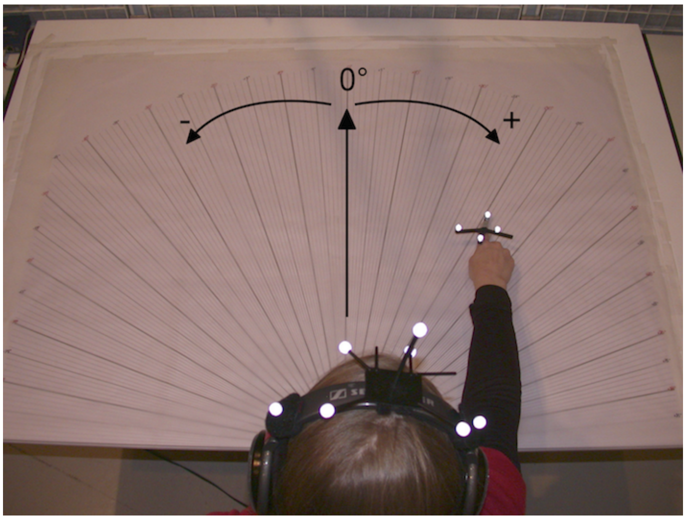


Figure 3.1 – View of the experimental set up, protractor on the table (0° axis straight ahead) and optical markers of the Optitrack 3-D motion capture system on the head (attached to the headphones) and right hand of the subject. Note the positive/negative angles reference.

3.2.3 Experimental procedure

The experiment lasted 1 hour and was composed of pre-trials and 4 sessions. The pre-trials aimed at selecting the best-fitting HRTF from a set of several HRTFs. This best-fitting HRTF was then used to convolve the stimuli of the main experiment. Subject tested HRTFs previously selected in HRTFs fitting past experiments (see (Sarlat et al., 2006) for a description of the method) plus their individual HRTFs when available, while hearing the spatialized targets. Up to four functions were tested. Approximately 10 practice trials per tested HRTF were performed in a pseudo-random order using the five targets of the experiment. Subjects were asked if they heard a spatialized sound and if so were asked to point towards its direction. The HRTFs were selected if in at least 8 trials the subjects pointed towards the correct direction (±10° approximately). The five subjects who did the pre-trials with their own HRTFs used them. The other subjects did not have individual HRTFs and used the non-individual HRTFs they selected during the pre-test.

Each session tested a different condition. In the short sound condition (named A) the auditory target was played for 250 ms before subjects pointed towards it. In the long sound condition (B) the auditory target was played for 2000 ms and subjects pointed towards it whilst hearing the auditory stimulus. Two other sessions included the auditory avatar that provided auditory feedback of the position of the hand in space. The fingertip position was dynamically tracked in real-time with the motion capture system and controlled the sound spatialization. Thus the white Gaussian noise stimulus was perceived as coming from the hand position. In these sessions the target was displayed during 250 ms and the avatar was heard constantly. In the “avatar condition” the actual hand position was heard (C), and in the “conflicting avatar condition” (D) the audio rendered hand position was shifted 18.5° left from the real hand position. Before each session, the subjects did a few trials to get used to the task demands and to the auditory feedback. The subjects were divided into 2 groups: group 1 performed the sessions in the regular order (A-B-C-D) and group 2 in the reverse order (D-C-B-A).

At the beginning of a trial, subjects were told to put their right hand on the table in front of them near their abdomen, with the palm at a position indicated by a tactile marker, and to hold their head up right facing ahead during the experiment. The auditory sources originated from a virtual distance of 60 cm in the horizontal plane of the table centered by the tactile marker. The targets originated from five directions with azimuth angles of −35°, −20°, 0° (ahead of the subject), 20°, and 35° (right is positive). Each session contained 32 trials presented in the same pseudo-random order for each subject. Moreover, the table on which the subjects pointed was covered with a semi-circular protractor of which origin was located at the starting hand position. It enabled a measure in degrees of the pointing as subjects were asked to keep their hand still for a few seconds after pointing. After each trial the subjects put their hand back to the tactile marker. The experimental setup from the subject’s viewpoint is shown in figure 3.1.

3.3 Data analysis

3.3.1 Level of performance

The pointing direction was directly measured on the protractor. The level of performance is evaluated by the signed angular error, which is the difference between the target direction and the final direction pointed by the subjects. If the subject pointed to the left of the target, the error was negative, and conversely it was positive if the subject pointed to the right of the target.

3.3.2 Movement analysis

The raw data of hand and head positions was recorded and processed off-line for the analysis of the kinematics of hand and head movements. A semi-automatic method was designed to detect and segment each pointing gesture and eliminate the way back to the start tactile marker. A primary segmentation was performed by applying thresholds on the hand displacement along the horizontal plane (x, y). The typical trajectories projected on the horizontal plane are shown in figure 3.2 for each condition.

The second segmentation process was based on systematic movement kinetics analysis. To compute velocity, acceleration and jerk, position data was filtered with a Gaussian low-pass filter, with a cut-off frequency of 5 Hz. As the movement is captured along the three dimensions of space the computed values are 3-dimensional energy-related vectors:

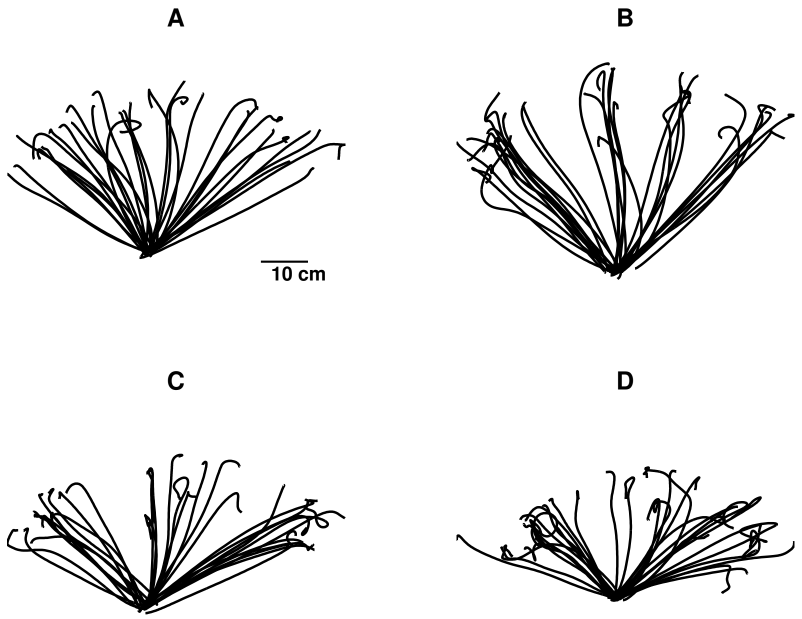


Figure 3.2 – Typical trajectories of the tracked hand for a single subject for each of the four conditions tested: short sound condition (A), long sound condition (B), avatar condition (C), and conflicting avatar condition (D). Better pointing precision and reduced overshooting is noticeable in condition (B).

v_{3D}, a_{3D} and j_{3D} are respectively the norms of the tangential velocity, acceleration and jerk vectors. The beginning and the end of movement were defined as the crossing of a threshold on v_{3D} corresponding to 3% of the peak velocity calculated on the trajectory. The "beginning" of the gesture is thus related to the energy of the movement. The typical velocity and acceleration profiles obtained for one pointing gesture are plotted on figure 3.3.

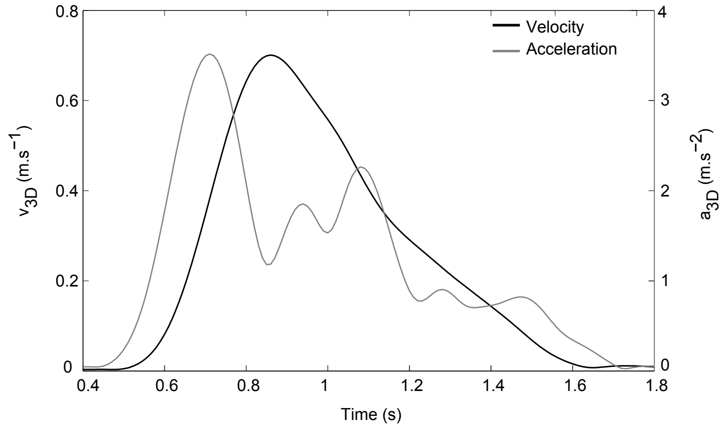


Figure 3.3 – Typical tangent velocity v_{3D} (bold line) and a_{3D} (gray line) profiles of a pointing movement.

Additionally, kinematic analysis included the following measures for hand and head movement: movement duration, peak velocity value, average velocity, acceleration peaks analysis (occurrence and position), and trajectory length in space. We counted the total number of acceleration peaks occurring before and after the maximum velocity peak of the movement (peak velocity point PVP). In order to investigate the possible role of the head in sound localization before and during pointing to the estimated location of the source, we also measured the heading angle around the vertical axis and computed its maximum values and range of motion (ROM).

3.4 Results

3.4.1 Statistical analysis

The results of 6 participants were removed from the analysis based on three criteria: subjects who did not follow the instruction to point directly towards the target (the trajectory duration is more than twice the average and longer than the longer stimulus duration in the long condition) - three subjects; trajectories showing no dependence on the target direction (with only two ± 90° endpoints) - two subjects; short trajectories (less than 10 cm) that lead to unstable angular calculations - one subject.

The dependent variables considered in our statistical analysis (ANOVA) are the averaged measures (duration, maximum velocity, average velocity etc.) over each target direction and each condition. In the statistical analysis, we considered two grouping factors. The first is the two-level HRTF factor that indicates if the subject used his own HRTF or not. The second factor is the two-level group factor that indicates the order of the presentation of the experimental conditions. We also considered two repeated-measure factors. The first one, the five-level target direction factor, corresponds to the direction of the target. The second one is the four-level condition factor indicating the experimental condition of each trial (A-B-C-D).

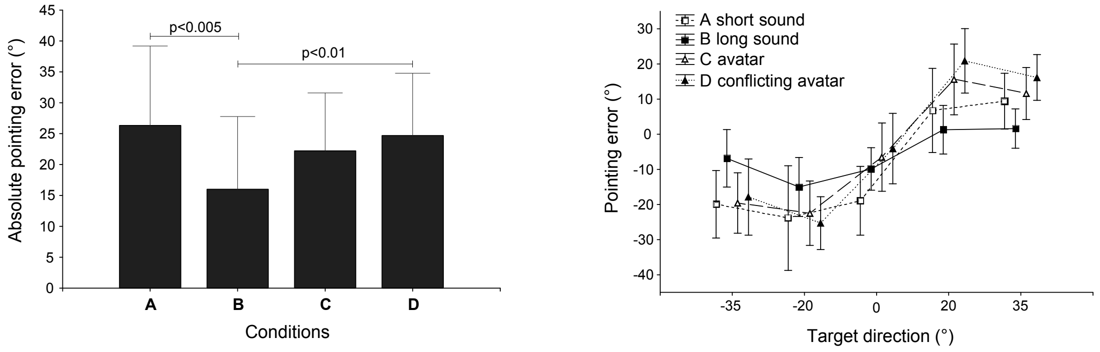
Statistical data analysis showed no main effect of the group factor. There was thus no effect of the order of the conditions either on the pointing performance or on the dynamical control of the gestures. There was a main effect of the individualized HRTF only on the proportion of acceleration peaks of the head after the PVP ($F(1, 16) = 5.8, p < 0.05$). However the average peak number was not significantly different between the two levels of the HRTF factor (post-hoc Bonferroni test). It is important to note that the individualized HRTF factor had no effect on the measures related to hand movement. The group factor and the individualized HRTF factor will not be used further in the analysis and data will be averaged per factor.

3.4.2 Level of performance

There was a main effect of the condition factor but also of the target factor on the absolute value of the angular error ($F(3, 51) = 6.23, p < 0.005$ and $F(4, 68) = 5.80, p < 0.001$

respectively). Subjects were significantly more accurate in the long sound condition B (see figure 3.4a which shows the absolute pointing error for the different conditions and the results of the post-hoc Bonferroni test - error bars indicate 95% confidence interval). Furthermore there was a significant interaction between the two factors ($F(12, 204) = 1.91, p < 0.05$).

We also analyzed the signed angular error as the sign indicates if the subjects pointed more to the left or more to the right of the target direction. There was a main effect of the condition ($F(3, 51) = 2.84, p < 0.05$) and the target direction factor ($F(4, 68) = 20.34, p < 0.0001$), and there was a significant interaction between condition and target direction factors ($F(12, 204) = 6.13, p < 0.0001$). Targets' azimuths were over-estimated by the subjects (see figure 3.4b which shows the signed pointing error for target directions and among conditions tested). Left targets were pointed with negative errors, and right targets with positive errors. This overshooting was reduced in the B condition: -66%, -40%, -48%, -94% and -90% for targets from left to right compared to the maximum errors in the other conditions. However, it is important to note that the subjects still presented a 9.8° average bias on the left when the target was presented straight ahead in the B condition.



(a) Pointing error *vs* conditions

(b) Pointing error *vs* target directions

Figure 3.4 – Absolute pointing error in degrees (absolute difference between pointed direction and target direction, (a)) for each condition and signed pointing error in degrees (difference between pointed direction and target direction, (b) for each target direction and each condition. Target direction goes from left (negative) to right (positive). A positive error indicates a pointing to the right of a target. Bars indicate 95% confidence interval.

3.4.3 Global kinematics

The parameters associated with movement velocity were significantly influenced only by the target direction ($F(4, 68) = 8.66, p = 0.00001$ for the duration ; $F(4, 68) = 81.59, p < 0.00001$ for the peak velocity and $F(4, 68) = 72.31, p < 0.00001$ for the average velocity). Peak and average velocities were significantly higher for target sounds coming from the right (*i.e.* for +20° and +35°): +37% for peak velocity and +31% for average velocity, post-hoc Bonferroni test $p < 0.0001$. The same test revealed no exploitable difference between the five target directions regarding movement duration. The condition factor, the target direction factor and their interaction had a significant effect on the trajectory length

($F(3,51) = 5.47, p < 0.005$; $F(4,68) = 47.03, p < 0.0001$ and $F(12,204) = 2.58, p < 0.005$ respectively). The analysis showed a significantly longer distance covered for targets on the right (0.473 m at +20°, 0.510 m at +35° against 0.414 m for the three other targets averaged, $p < 0.005$), but also in the B condition (0.482 m against 0.432 m on average, post-hoc Bonferroni test $p < 0.05$; see figure 3.5).

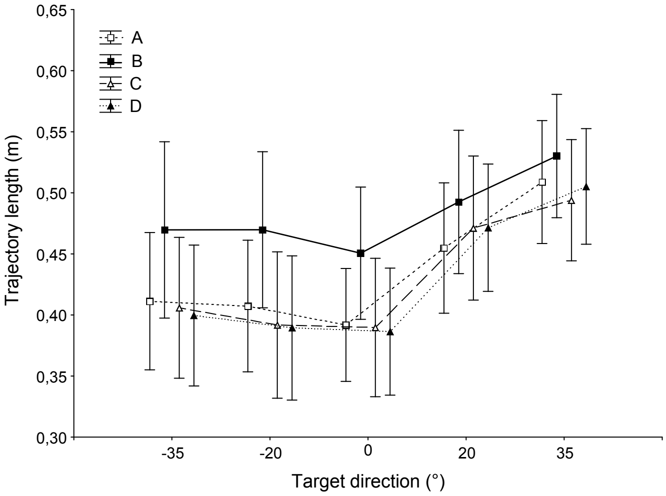


Figure 3.5 – Trajectory length in space (in meters) for each target direction and each condition. Bars indicate 95% confidence interval.

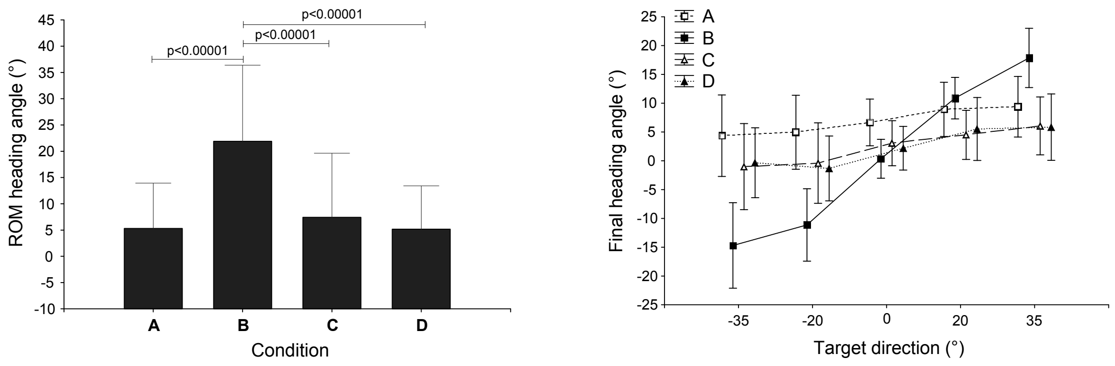
3.4.4 Movement dynamics and segmentation

The counting of acceleration peaks revealed a significant effect of condition factors ($F(3,51) = 3.04, p < 0.05$) and target direction ($F(4,68) = 30.93, p < 0.00001$) on the total number of peaks and on the proportion of peaks before reaching the PVP ($F(3,51) = 3.34, p < 0.05$ and $F(4,68) = 36.97, p < 0.00001$). In the B condition, subjects' movements presented larger total number of acceleration peaks, however not significantly different from the other conditions (4.85 against 4.20 on average). The number of peaks decreased as the target direction shifted to the right of the subjects (significantly for the two targets on the right, post-hoc Bonferroni test $p < 0.0005$: 3.89 and 3.70 against 4.73 on average). Only the target direction factor had an effect on the proportion of peaks after PVP ($F(4,68) = 10.68, p < 0.00001$) significantly different for +20° and +35° targets (-18% at +20° and -22% at +35° on average), while there was a marginally significant effect of the condition factor ($F(3,51) = 2.47, p = 0.07$).

It is noticeable that subjects produced movements with more acceleration peaks on the second "half" of the trajectory, during the deceleration phase: 1.52 before the PVP, 2.84 after on average. If taken as a factor, the proportion of peaks before or after PVP together with the condition factor shows a significantly higher increase of peaks after PVP for condition B than conditions A and C (post-hoc Bonferroni test $p < 0.01$).

3.4.5 Head movement analysis

The same analysis was conducted on the head movement data. The target direction factor had a significant effect on the total number of acceleration peaks in the head movement ($F(4, 68) = 5.75, p < 0.005$) with the same tendency towards right directions as for the hand (7.14 peaks for -35°, 6.53 for +35°). No significant effect was found on the proportion of acceleration peaks before PVP. After this point, both target direction and condition factors have significant effects ($F(4, 68) = 4.97, p < 0,005$ and $F(3, 51) = 6.93, p < 0.001$ respectively), again with the same behaviour as for the hand. The B condition exhibited a significantly larger numbers of peaks after PVP (+50% for B on average, post-hoc Bonferroni test $p < 0.05$) than in the other conditions and the centre and right targets exhibited fewer acceleration peaks (-17% on average). Both condition and target direction factors had a significant effect on the ROM of the heading angle ($F(3, 51) = 20.2, p < 0.0001$; $F(4, 68) = 3.93, p < 0.01$ respectively) and there is a significant interaction between the two factors ($F(12, 204) = 2.40, p < 0.01$). The ROM of the heading angle was significantly higher in the B condition than in the other conditions (21.9° against 5.31°, 7.42° and 5.17° for A, C and D conditions, post-hoc Bonferroni test), as shown in figure 3.6a. No significant difference was found among the target directions but the ROM increased with the target eccentricity (+45% on the left, +23% on the right on average compared to 0° target).



(a) ROM of heading

(b) Final heading angle

Figure 3.6 – Range of motion of the heading angle (in degrees) for each condition (a). Effect significance: ($F = 20.2, p < 0.0001$); post-hoc Bonferroni test; Final heading angle (in degrees) for each target direction and each condition (b). Interaction effect significance: $F = 14.9, p < 0.00001$. Bars indicate 95% confidence interval.

In order to investigate the potential link between target direction and head rotation for localization when pointing we analyzed the distribution of the heading angles at the end of the movement. As for the ROM, the condition factor, target factor and their interaction had an effect on the angle ($F(3, 51) = 5.07, p < 0.005$; $F(4, 68) = 9.17, p = 0.00001$ and $F(12, 204) = 14.9, p < 0.00001$ respectively). Significant differences were found for the two right targets compared to left targets ($p < 0.01$); the subjects turned their head towards the correct hemisphere corresponding to the target direction. When coupling the effect of the condition and the target direction, we found that this behaviour was prevailing under condition B (see figure 3.6b). The two graphs on figure 3.6 show that subjects moved their head more under condition B and in the direction of the target. The bias for 0° target is also reduced under this condition: 0.30° compared to 6.66° for A, 3.06° for C and 2.18° for D.

The analysis of the relative position of the PVP along the movement of the hand and the heading angle shows that subjects tended to initiate the movement of their head before the pointing movement. The distribution of these relative positions is shown in figure 3.7 for every trial over every subject in each condition. On average, 43% of the gestures exhibited heading peak velocity between the beginning and the first third of the movement completion against 12% only for the hand. The tendency is observed in all the conditions and in spite of the large differences in ROM of heading and final angle between conditions.

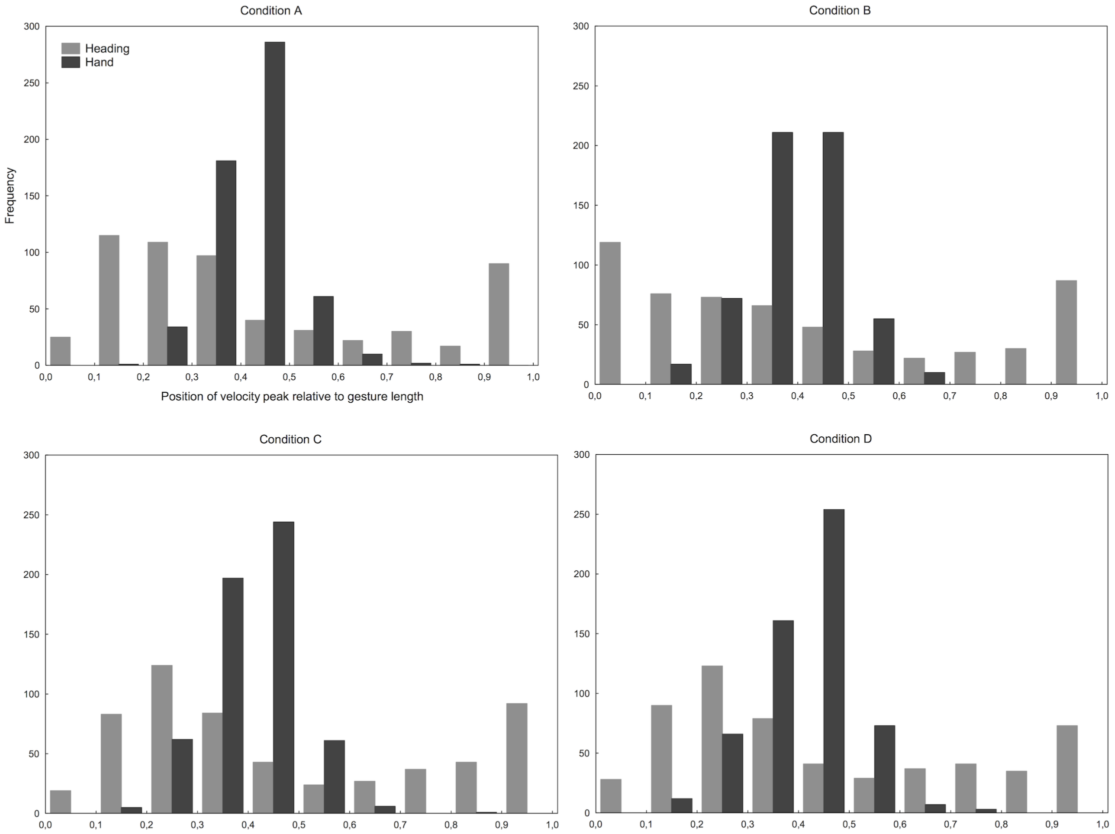


Figure 3.7 – Distribution of the position of the peak velocity point of heading angle and hand relatively to gestures length for every trial over every subject in each condition: short sound condition (A), long sound condition (B), avatar condition (C), and conflicting avatar condition (D). It indicates that head maximum velocity is reached sooner than for the hand.

3.5 Discussion and conclusion

In this study, we attempted to address the mechanisms whereby an auditory input is transformed into a motor command. First, we aimed at assessing the role of auditory information about target position in correcting the trajectory of the hand by varying the duration of the target presentation. Second, we attempted to evaluate whether human subjects could use an auditory feedback about their hand position and how they would react to shifts in this avatar of their heard hand position.

Only the long sound target condition exhibited a higher level of performance of the subjects. This strong effect is comparable to the one obtained during pointing movements towards visual targets present throughout the entire pointing movement (Prablanc et al., 1986). In the present study, the target is presented during the whole movement only in the long sound duration condition (B). In the short sound duration condition, the location of the target needs to be memorized and it is possible that a shorter sound would lead to a less precise or reliable representation of the target. Errors in pointing to remembered targets presented visually have been shown to depend on delay between target offset and pointing (McIntyre et al., 1998). Therefore, the neural processes involved in coding the target in a motor-related or body-related reference frame from its auditory spatial trace seem to require a sufficiently long auditory stimulation. On the other hand, one can assume that comparison of auditory information about target position with proprioceptive information is required to update or refresh an internal representation of the goal to drive optimally the pointing hand.

In addition to better performance and precision (reduced bias for 0° target), subjects presented longer trajectories in the longer sound condition and slightly more acceleration peaks. The proportion of acceleration peaks in the deceleration part of the movement also increased in this condition. These results show that the improvement of precision in this condition may not only be due to better memorisation of the target but also to the possibility to make online corrections of the hand trajectory. The use of auditory information about target direction as a feedback for guiding the reaching movement is likely since the kinematics showed indices of iterative corrections in condition B (in particular, increased length of the trajectory and increased number of peaks after PVP). These online corrections can be produced only if a neural process is able to use the auditory estimation of the target position and to make it available continuously to the sensorimotor process that drives the hand. Therefore, a sound still heard at the end of the pointing movement as in condition B would allow a more efficient updating of the goal representation in relation to the hand's position and thus a more accurate movement.

Contribution of the auditory avatar

As demonstrated in Oscari et al., 2012, hand trajectory can be controlled and optimized with an auditory feedback. Here, the directional accuracy of pointing movement was not greater with auditory feedback of the hand position than without this information available (comparison of conditions A and C). Furthermore, in condition D auditory feedback of hand position was shifted by 18,5° perpendicularly to the main movement direction. Following the shift, the hand trajectory was expected to deviate from those produced in the condition without the shift. The analysis showed no significant effect of the resulting discrepancy between auditory and proprioceptive information about hand position on the pointing accuracy. It is possible that the levels of performance in all conditions but the long target condition were impeded by an inaccurate representation of the target relative to the body and that this important inaccuracy masks a small effect of the hand auditory feedback. Indeed, in the short sound condition with no avatar (A), the mean absolute pointing error was of 26°, higher than the shift used with the avatar in condition D. In the avatar conditions, the proprioceptive modality also might have overtaken or dominated the overflowed auditory modality, hence the importance of the design of such feedback, as showed in (Rosati et al., 2012). In their study, the authors compare the contribution of different sound feedbacks on the performance in a manual tracking task and their interaction with visual feedback. They have observed that sound feedback can be counterproductive depending on the task and mapping between gesture and sound. In our experiment the same sound was used for the targets and the hand feedback. This might have confused the subjects when localizing the target and addresses the question whether spatial auditory information about limb position is enough to provide an efficient feedback to a motor action. Different parameters of the motor action might indeed need to be sonified (for instance kinematics rather than position in space). It is therefore important to study the appropriate parameters for auditory-motor mapping before being able to provide useful information for rehabilitation and sensory substitution devices.

Head movements

The analysis of final head orientation showed that in B condition heading automatically accompanied the auditory-manual pointing task despite the explicit instruction to avoid head movements. Thus, head rotations were only present when sufficient localization cues were available and the heading direction was consistently related to target direction and eccentricity. The first hypothesis than can be proposed is that this result indicates that in all the other conditions tested, the auditory target was too short to provide enough information to elicit head movements. However, since the heading direction and the direction of the pointing are clearly related in condition B (see figure 3.6), one can propose also that the long sound allows an orienting movement of the head towards the auditory target and that the final angle of this orienting movement could guide the pointing movement of the hand. The fact that the head tends to achieve its maximum heading velocity before the hand PVP in all the conditions (see figure 3.7) shows that early movement of the head alone did not lead to improved performance in condition B, but did along with a larger ROM and heading towards the target.

In general, heading movements belong to automatic orienting reactions that have been mainly studied in the framework of gaze orienting behaviour (Guitton, 1992). Here in blindfolded subjects, we can assume that heading also aims at optimizing the binaural perception of the acoustic stimulation direction. The auditory system certainly relies on head motor information to build representations of the location of auditory targets. However and unfortunately, sound localization is mainly studied with the head fixed. Nevertheless several studies have used head orientation to quantify the ability of participants to indicate the perceived direction of a natural acoustic stimulation (Makous and Middlebrooks, 1990; Perrott et al., 1987; Pinek and Brouchon, 1992). These studies demonstrated that the direction indicated by the head was underestimated ($\sim 10°$). We obtained similar results despite different experimental conditions (voluntary head pointing versus automatic orienting reaction). Orienting reaction and voluntary heading to natural acoustic stimulation were observed with relatively short stimuli (500 ms) in (Goossens and Van Opstal, 1999). In contrast, in our experiment with HRTF spatial rendering, heading toward the target was little observed with short sound stimuli. However, Goosens and Van Opstal suggested that head movements could provide spatial information about rich and long enough sounds that would be used by the auditory system to update the internal representation of the target. Our results suggest indeed that the accuracy of pointing to long stimuli could be due to the contribution of heading toward the target providing a more accurate frame of reference for the anticipated control of pointing. However, this does not exclude a direct role of the on-going presentation of the acoustic target.

Target directions and characteristics of movements

The estimated direction of targets are characterized by a perceived space wider than the real one. This was also observed with hand pointing toward “natural” sounds produced by loudspeakers (Pinek and Brouchon, 1992). However it was much larger in our study than that observed with natural sounds (less than 10° for Pinek and Brouchon) and this could originate in the use of non-individual HRTF in which interaural differences are not adapted to the geometry of the head. The observed left bias in direction for straight ahead targets could result from a pseudo-neglect effect favouring the left hemispace similar to the pseudo-neglect effect observed with vision (Sosa et al., 2010). The left/right asymmetry observed in the trajectories kinematics can be explained by this effect as well. Indeed average and peak velocities increased for targets on the right without effect of the conditions. Along with longer distances covered and fewer number of acceleration peaks, this effect might have caused variations in the control parameters of the movements between the two hemispaces. The left bias observed for the 0° target sound supports this hypothesis. Nevertheless, considering the starting position of the task with the palm put at the centre of the set-up, these results could also be accounted for subject’s ease to point on the right with their right hand.

Modularity

This study addresses also the question of the cooperation between different modular neural processes involved in the multisensory and motor representations of targets in goal-directed movements. Do these different processes share a global amodal spatial representation (*e.g.* (Pouget et al., 2002)) or do they have their own dedicated spatial representation? Visual and auditory modules use certainly very different reference frames. Sounds are localized thanks to spectral and binaural cues naturally linked to a head-centred frame of reference when visual positions are primarily coded in an eye-centred reference frame. In addition, the visual system is retinotopic whereas the auditory system is characterized by broad tuning and lack of topographical organization (Maier and Groh, 2009).

The question of modularity in motor control arises when we consider the coordination between head orienting movements and hand movements. In the longer sound condition, the auditory stimulation is long enough to allow the triggering of head rotations. Since the amount of rotation of the head is related to the response of participants, there should certainly be a way for the two processes to share common information. This suggests that the heading direction is coded in a body-centred reference frame and can be used directly by the reaching motor command that shares the same reference frame.

To conclude, it is known that sound localization requires the integration of multisensory information and processing of self-generated movements: a stable representation of an auditory source has to be based on acoustic inputs and their relation to motor states (Aytekin et al., 2008). Our results highlight that auditory representations extracted from a sound signal can be transformed online into a sequence of motor commands for coordinated action, underlying the role of the auditory-motor loop in spatial processing.

Conflict of interest statement This research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgments We wish to thank the reviewers for their constructive comments on previous versions of the manuscript. This work was supported in part by the ANR (French National Research Agency) Legos project (11 BS02 012) and by the CNRS MI (Mission Interdisciplinarité) DEFI-SENS. It has been accomplished within the laboratory of Excellence SMART supported by French state funds managed by the ANR within the Investissements d'Avenir program under reference ANR-11-IDEX-0004-02.

Chapter 4

Touching sounds: gestural interaction with a virtual sonified object

Abstract: This chapter presents a study exploring the role of interactive movement sonification in the case of a non-tangible (virtual) object. The interaction with the object is motivated by sensing its geometrical properties with the hand. Sound is used to 'create' this object through the gestural sonic interaction, conceptualized in an "Auditory Virtual Surface". The framework we developed allows to investigate the gesture-sound relationships under a sensory substitution paradigm, as visual and tactile inputs are suppressed. This study addresses the following question: is it possible to perceive a geometric property of a virtual object only through a gestural sonic interaction? Subjects are asked to determine the curvature direction of a (virtual) thick plate with the hand. We investigate the role of sonification in this task with two different cases, static and static+dynamic. An evaluation of the perception capabilities of the subjects is performed, and gestural features are examined. Results show that the gesture-sound interaction allowed the subjects to perceive the curvature of the plate. The dynamic sonification proved to be more efficient in this task and different gestural strategies of exploration emerged. Observations confirms that movement sonification can be used to create a gestural interaction with a virtual object, leading to adapted perceptual abilities. This supports the use of the developed framework and the potential of movement sonification in sensory substitution. Although sound is currently used in sensory substitution devices (especially for visually impaired people) we argue that an interactive context would benefit perception, learning and engagement.

Selected results will be presented in March 2015 at the VR 2015 IEEE Virtual Reality conference. Publication version in preparation.

4.1 The sensory substitution paradigm

Sensory substitution is a perceptual process where a particular sensory information is delivered through either another sensory channel (usually not perceiving that information) or the same channel but with a different body part using this modality. For example, Braille is a sight-to-touch sensory substitution system allowing to perceive through touch (the tactile pins on the paper) information usually acquired through vision (graphic symbols representing words).

The earliest work we found traces of dates back to 1897, investigating the utilization of touch and skin receptors instead of vision. But already in ancient times the mysteries of sensory illusions questioned the philosophers. The Aristotle illusion, scientifically named (but less poetically) *diplesthesia*, illustrates the possible discrepancy between two senses and presages of sensory substitution. Aristotle (384–322bc) noticed that when one crosses two fingers, an object he touches would feel like two objects (see an illustration figure 4.1). To quote exactly *On Dreams in the Parva Naturalia*:

> "When the fingers are crossed, one object seems to be two; but yet we deny that it is two; for sight is more authoritative than touch."

This observation pointed out the dominant confidence in sight and the subjectivity of visual perception, a kind of challenge for sensory substitution.

Figure 4.1 – Aristotle illusion. One would feel two objects when crossing fingers.

The first approach of sensory substitution focused on finding an alternative way for visually impaired people to perceive visual information from the world. Back in the late sixties, Bach-y-Rita presented the first experimental results of a vision-to-touch sensory substitution device (Bach-y Rita et al., 1969; Bach-y Rita, 1970). Visual information is captured by a TV camera, spatially coded, and transmitted to a matrix of 400 vibro-tactile actuators of 1 mm diameter placed on the back of the subject (Collins, 1970) (the curious reader is invited to look at the complex, yet ground-breaking experimental setup the authors developed to this purpose). The spatial resolution of the tactile image reproduced on the subject's back is approximately one tenth of the resolution of the fovea (the most sensitive part of the retina in human eye). Experimental results show that subjects were able to perceive the shapes of common objects or faces, and even moving stimuli. Interestingly, the blind subjects interpreted these stimuli rather as visual than cutaneous. Bach-y-Rita writes in his 1970 paper: “(...) it would appear that the central nervous system is capable either of utilizing existing mechanisms or of developing new mechanisms to process the information”.

The main mechanism that allows for such an (re)interpretation of sensory input is precisely this ability of the central nervous system to adapt and modify its network (specifically the perceptual circuits here) responding to a modification of the incoming information circuit.

4.1.1 Neural basis: cerebral plasticity

To give a quick and simple definition of cerebral plasticity, we chose the definition Bach-y-Rita quotes in (Bach-y Rita and W. Kercel, 2003): “the adaptive capacities of the central nervous system – its ability to modify its own structural organization and functioning”. It is driven by functional requirements and rests upon changes at various levels in the brain (chemical, synaptic and neuronal). The plasticity can be temporary as it is not thought to change the fundamental functional representations in the brain (Jenkins et al., 1990). It is also a fundamental basis of learning mechanisms and stands as a response to training (see chapter 2), this is why it is the main targeted mechanism of rehabilitation processes (Krakauer, 2006) (see also section 2.3.4).

4.1.2 Experimental tool and compensatory apparatus

The feelings described by the blind patients in the study of Bach-y-Rita illustrate the presence of synaesthesia in sensory substitution (O’Regan, 2011; Proulx, 2010). The visual feelings came from a different modality than the actual modality (tactile) they used to perceive the information. Synaesthesia can be acquired or synthetic (Proulx, 2010) and can occur after intensive training on a sensory substitution system. A substantial number of studies on perception used the sensory substitution paradigm to investigate synaesthesia (Ward and Meijer, 2010). It is interesting to note that during synaesthesia both the transformed and the original sensations exist side by side, so that people can consciously build an analogical relationship between the two (like ‘seeing’ words or numbers in shapes and colors (Tammet, 2006)).

The most successful and well-known sensory substitution system is Braille. Numerous systems have been developed, mainly vision-to-touch or vision-to-auditory, but also auditory-to-vision (e.g. enhanced visual displays for deaf people). In 2003 Kercell and Bach-y-Rita used the tongue and electrical stimulation of this muscle to translate visual information from a camera placed on the head of the subject (Bach-y Rita and W. Kercel, 2003). Similarly to the vibro-tactile actuators they have placed on the patient’s back, they used the extreme sensitivity of the tongue, the protected environment of the mouth and the presence of the saliva which ensure a good electrical contact. They also pointed out that with an equivalent electrical power the tongue was more sensitive than the fingertips. They placed a 12x12 matrix of electrodes that delivered bursts of electrical pulses through the tongue, encoding the image of a CCD camera. Subjects were able to perceive the electrical information as an image in space, although after a period of training. They were also able to make perceptual assessments using visual concept (shapes, contours, perspective), again a form of synaesthesia.

4.2 Audio-based sensory substitution

The auditory modality presents a number of advantages for sensory substitution. It offers a perceptual channel generally available with a limited cognitive cost, can operate in the dark, and does not interfere with body movements and mobility. However, the localization of sound sources by the human hearing system suffers from a lack of precision (Makous and Middlebrooks, 1990) and biases (Dufour et al., 2007). The use of spatialized sound to convey spatial information about objects or movements is indeed still debated (see chapter 3 (Boyer et al., 2013)) (Parseihian, 2012). Nonetheless, a large number of auditory sensory substitution devices are available today. It represents an important field of the research in rehabilitation and impairment devices. Besides, and especially today, audio is available at low cost and on mobile platforms with a decent quality of restitution.

We describe quickly here three auditory sensory substitution devices that were developed by research institutions to help visually impaired people. The Voice, developed in 1992 by Meijer (Auvray et al., 2007) uses a head-mounted camera and translates gray-levels images into sound by a simple sine wave-based sonification process. The scene is described laterally from left to right, so the position in time of the elements heard in the sound give their azimuthal position. The pitch of the sounds are mapped to their vertical position in the image. Finally, the intensity corresponds to the gray level in the image. This mapping, although basic, requires a long learning time to use the system. Yet Ward and Meijer confirmed that subjects were able to learn and acquire a synthetic form of synaesthesia (Ward and Meijer, 2010).

Similarly, the Vibe system (Hanneton et al., 2010) encodes the image with multiple and configurable units to multiple sound sources, using a retina-like scheme of receptive-fields. The camera can be held in the hand of the subject so that movement is needed for the exploration, which is more active. An experimental validation showed that after a relatively short time of practice, subjects were able to achieve a pointing task when they held the camera in their dominant hand.

Finally, Bologna et al. (Bologna et al., 2009) experimented the sonification of colors in a head-mounted camera stream, where a HSL color coding system (hue, saturation and luminosity) is translated into different instrument timbers, notes and gains. The experimental tests included a colored socks pairing test and navigation following a colored line painted on the ground. They attempted to make the subjects explore colorful images and describe them but the auditory mapping turned too complex to interpret.

4.3 A need for interaction and movement

The various systems described previously do not necessarily engage specific actions by the subject. We believe that these systems could benefit from a positive interaction rather than simply make the subject learn to use the new 'sense' created by the sensory substitution and synaethesia feelings. As shown is chapter 2, sound is an excellent candidate to deliver fast, precise, metaphorical and even spatialized information, providing careful interaction and sound designs.

Sensing precisely the size of an object without seeing or touching it appears difficult with current sensory substitution devices. Touch is a *proximal* sense whereas audition is a *distal* sense as the localization of an auditory stimulus in space has to be extracted from the information that is perceived. Tactile (and haptic) perception implies contact, the production of mechanical forces and the perception of interaction forces, whereas audition consists in sensing changes in the air pressure field. Using an audio-visual interactive display, Avanzini et al. emphasized that auditory feedback plays a role in the perception of size and weight of objects (Avanzini et al., 2004). We propose to use interactive sonification within the sensory substitution paradigm to sens objects and explore their physical properties.

4.4 The concept of Auditory Virtual Surface

We define the concept of “Auditory Virtual Surface” (AVS) as an interactive region in space activated when the user's arm enters or moves into it. The three-dimensional coordinates and speed of the user's hand are used to synthesize real-time auditory feedback. The term 'surface' stands for interacting with a region that actually has a thickness, as a real solid surface could not be passed through. The virtual surface can thus have a thickness in space and be an interactive object throughout the thickness. For instance, in the setup developed in the following parts, the surface reacts as a physical one with a clear separation between the outside and the inside, but the sonic interaction can take place wherever in the depth of the region. We hypothesize that this concept allows for investigating the role of the auditory-motor coupling in exploring virtual interactive objects.

4.5 Questions

The following experiment addresses the sensory substitution of touch by sound. The general question we address is the following: to what extent is it possible to use sound to substitute from touch? Specifically, we investigate whether it is possible to perceive a geometric property of an object using auditory sensory substitution and the concept of AVS.

4.6 Experiment: sensing a geometric property of a virtual sounding object

We present here an experimental design to explore the auditory-motor coupling in a virtual sound-object relationship. We apply the previously defined concept of auditory virtual surface to make blindfolded participants sense a virtual object using exclusively the auditory modality. The aim of the procedure is to determine a geometric property of the object, namely the curvature direction of a thick plate (concave or convex). The virtual auditory relationship is investigated comparing a static and a dynamic auditory feedback. First we demonstrate that participants are actually able to imagine the object and can interact with it, and to determine its geometric property. Second we show the benefits of the dynamic auditory feedback on the accuracy of the participant's perception threshold. Finally, through the analysis of the movement produced while exploring the AVS, we characterize two auditory-motor strategies that the subjects adopted. The use of continuous movement sonification along with a low-cost motion sensing device is particularly innovative in the auditory sensory substitution field. This experiment can stand as a 'proof-of-concept' for the usage of sound as a media in a virtual object relationship and towards a new framework for studying sensory substitution and auditory-motor transformations.

4.6.1 Participants

A total of twenty subjects, aged between 20 and 27, participated in the experiment (55% female - 45% male). They all self-reported that they had neither hearing nor physiological disorder. All gave written informed consent for their participation. The subjects were randomly assigned to two groups *G1* and *G2* receiving a different auditory feedback. The first group interacted with the AVS through a static position-based feedback indicating that the hand is "in" the AVS. The second received an additional feedback based on the velocity of the hand tangential to the surface.

All participants were blindfolded during the experiment, in order to avoid any visual feedback from their hand position in space, which they could use to build a visual representation of the plate. By suppressing sight, we focus the attention of the participants on proprioception and potential tactile sensations through the motion-sound interaction, encouraging them to move during the task and to be more active.

4.6.2 Methods

Experimental setup

The experiment was carried out in a double-walled sound-insulated booth used for audiometric measurements. The subjects stood in front of a LeapMotion™ device attached to the platform of a camera stand. The auditory feedback was delivered by two loudspeakers placed on the ground near the stand and tilted (of approximately 40°) in the direction of the subjects. The LeapMotion™ [1] device allowed for the motion sensing of the participant's hand by measuring the three-dimensional position of the center of the palm at a minimum frame rate of 100 Hz. It uses infrared beams and a pair of infrared cameras collecting the reflections of the light from the hands above its surface. The average latency given by the API of the device was around 3 ms. The data acquisition and the sound generation were processed by a dedicated program built under the Max/MSP[2] environment. The data were received in Max/MSP using a custom built Max object based on the LeapMotion™ API.

The subjects were instructed to stand and to place their preferred hand above the device in order to explore the space in front of them. The figure 4.2 shows a picture of a subject in place and a schematic of the AVS used in the experiment. The AVS was a cylinder section with a curvature radius R, a constant thickness ($e = 80$ mm) and width (1000 mm, adapted to the horizontal operating range of the tracking device). A positive value of R makes the plate concave and a negative one makes it convex, with the same curvature and centered in the same spot. In the z direction, the depth of the surface was also defined by the maximum range offered by the device (the limit of the plate was the limit of the tracking area, which was superior to the arm length). It was presented as a "curved plate with a constant thickness" to the blindfolded subjects. The height H of the surface center was adjusted to the natural initial resting position of the subjects (with the hand kept horizontally); it was generally around 30 cm.

Protocol: modified psychophysical staircase procedure

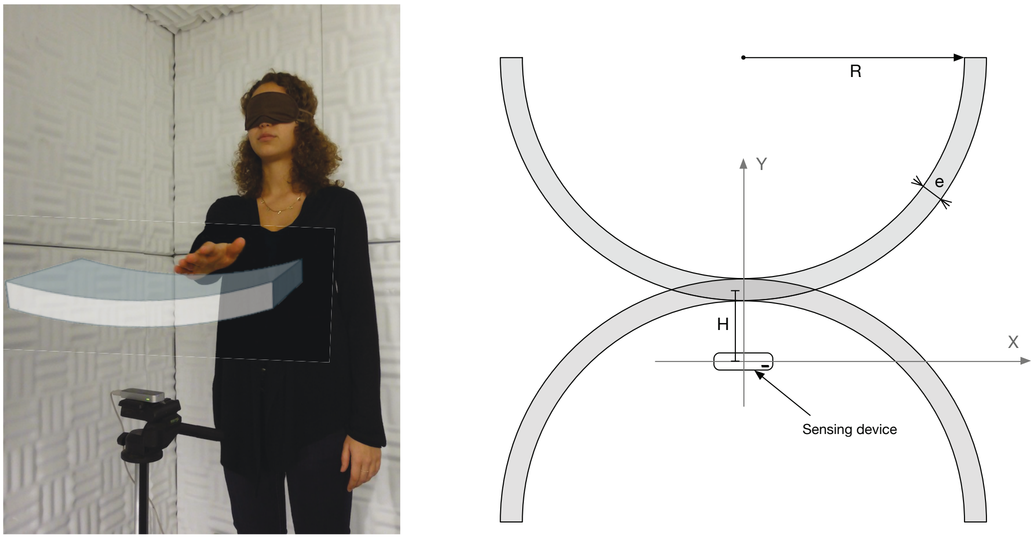
The subjects performed a discrimination task by answering the question "is the plate concave or convex?". They could actually answer with a set of words of their own choice if they were uncomfortable with these notions, after agreeing with the experimenter. In each trial, the blindfolded subjects were asked to explore the space with their hand and to determine if the plate was concave or convex (forced choice). If no answer had been given after thirty seconds, the subjects were asked to pick their choice as soon as possible. Once they gave their answer they stopped exploring the AVS and could rest by placing their arm naturally down along their body.

As the purpose of the experiment is to demonstrate the ability of subjects to discriminate

---

[1] `leapmotion.com`

[2] `www.cycling74.com`



(a) (b)

Figure 4.2 – Left: photo of a subject with an illustration of the auditory virtual surface geometry and the LeapMotion™ device. Right: schematic of the AVS. The stimulus of the experiment is controlled by the curvature radius R (value and sign) so the plate can be either concave (top one) or convex (bottom one).

between concave and convex surfaces, we used a simple staircase procedure (inspired from the spirit of psychophysics experiments (Levitt, 1971; Jesteadt, 1980) although in a very different manner). Subjects started with an initial curvature radius R of $R_0 = 400$ mm. After each trial R is changed depending on the previous answer of the subject. A block of ten random positive or negative values of R (concave and convex plate respectively) is used for each curvature value. The blocks were different for all the subjects. If the subject did not give three consecutive correct answers in the ten trials, then the experiment stopped and the current value of R was taken as the perception threshold of the subject. If he did, then R was increased (the plate becomes flatter) with a constant ΔR value and another block started. A first session with $\Delta R = 100$ mm gave a rough value of the curvature perception threshold R_{100}. To refine this value, a second session immediately started with slightly lower initial R ($R = R_{100} - 50$ mm) and step size $\Delta R = 50$ mm. We consider the final R_{50} value as our best estimate of the discrimination threshold of the subject.

By the end of the experiment, a short questionnaire was submitted to the subjects to evaluate their feelings about sight deprivation and their personal experience with the sonic virtual interaction.

Auditory feedback

The auditory interaction of being in the AVS (static position feedback) was designed as a synthetic sound texture using granular synthesis. The synthesis was programmed with the MuBu modules (Schnell et al., 2009), a series of Max/MSP objects built for interactive real-time audio processing. The granular synthesis techniques allows for playing overlapping short 'sound grains' from recorded sound samples. We used 211 ± 30 ms long grains, played every 32 ms period, from a sound sample of oscillators and filters. This method can create a rich sonic texture statistically stationary (so that the sound is 'the same' for all the subjects) but not constant to minimize lassitude. The sound is heard when the hand of the subject enters the surface limits and is turned off when it goes out. Linear ramps of 20 ms duration are used to smooth the triggering.

We chose to use a generic sound but keeping an ecological and understandable mapping (Gaver, 1993b; Gaver, 1993a) so the subjects could access it. The relationship between the presence of the hand in the AVS and the position feedback is only imaged. As this action is physically impossible (either you hit a solid surface, either you are out of it, but you cannot go through it), the information does not have an ecological twin. There is then no need for a particularly 'realistic' sound model. A physical model of a hit and scratched plate has been tested and revealed quite disturbing indeed. The reader should note here that we will mention a possible future development of this idea in the last part of this chapter.

The sound was generated from the hand position and velocity data as well as from the plate geometry. The custom program in Max/MSP included the design of the AVS (choice of geometric properties), data acquisition from the LeapMotion™ device, recording, sound feedback generation and rendering and experimental procedure control. Whenever the hand of the subjects entered the plate (*i.e* matched the equation of the cylinder section) the static sound feedback was heard. Subject assigned to the G1 group received only this position feedback.

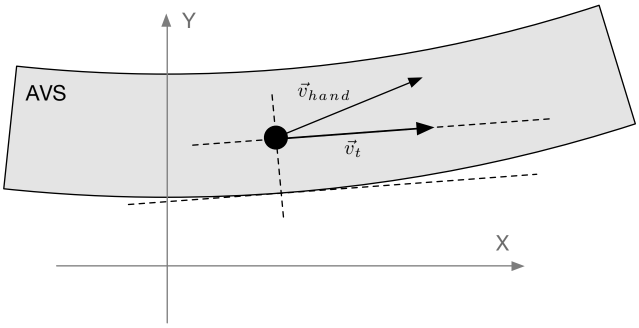


Figure 4.3 – Schematic illustrating the velocity vector of the hand $\vec{v}_{hand}$ and its component parallel to the AVS $\vec{v}_t$ which is sonified in group G2. Both G1 and G2 subjects receive the static feedback when the hand enters the AVS area.

For subjects in the G2 group, a dynamic friction-like sound was added to the static feedback by controlling the cutoff frequency of a resonant filter (1 dB gain, 0.7 q-factor) processing a pink noise (constant power per octave). The cutoff frequency is controlled by the velocity component $\vec{v}_t$ of the center of the palm parallel to the surface (figure 4.3). A 0-600 mm/s velocity is mapped on 400-5000 Hz frequency in the filter. This dynamic feedback is also turned off when the hand exits the surface with the same linear ramps as the static one. Figure 4.4 summarizes schematically the two sonification processes.

A stereo panning using the x horizontal position of the hand in the surface was added to move the sound laterally so it was spatially coherent with the hand position. To ensure an immersed interaction, the two loudspeakers on the ground were tilted towards the subjects so that the sounds were actually coming from underneath the subjects' hand.

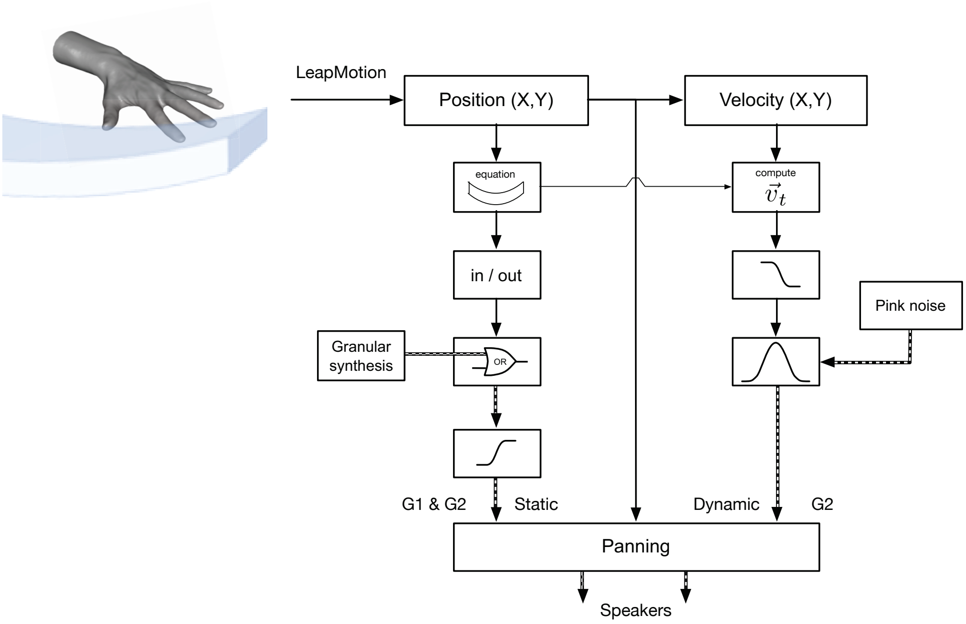


Figure 4.4 – Schematic of the static and dynamic sonification processes.

4.6.3 Data analysis

Since the first aim of the study is to demonstrate that the subjects were able to learn to discriminate concave from convex curvatures of the plate, we collected the estimated discrimination threshold in curvature units, that is $1/R_{50}$. We first tested the hypothesis that the difference with the initial curvature $1/R_0$ was due to chance with a Student-Fisher t-test. We then examined the following features in each trial and for both groups:

- curvature threshold $1/R_{50}$
- number of correct answers
- response time RT
- time spent in the AVS
- direction of velocity vector in the AVS

The response times (named 'RT', not to be confused with 'reaction time', often measured in experimental psychology and perception studies) were collected although their precision is quite low since subjects took some time to give their answer (in spite of the instruction to give a quick answer after thirty seconds of trial). The trajectories of the subjects' hand were analyzed for each trial searching for characteristic gestures and features of the movement performed while exploring the AVS. The time in contact with the AVS was expressed as a percentage of the total length of the exploration, which was measured as the time between the first contact and the last time subjects exited the plate.

4.6.4 Results

The first observation is that a large majority of the participants (90%) were actually able to perceive the AVS and to evaluate its curvature. Two subjects, one in each group, have been excluded from the threshold analysis due to their outstanding performances in the task. They reached 0.77 m^{-1} for the one in group G1 and 0.62 m^{-1} for the one from G2, that is 1300 mm and 1600 mm curvature radii respectively. At these very low curvatures, the thickness of the plate (80 mm) is twice as big as the declination to detect at the extremities of the operating area. Their performance overtake the average curvature threshold by a factor between two and three. Their case will be addressed in the Discussion paragraph.

Curvature discrimination threshold

The first group (G1) obtained a curvature radius threshold of R_{50}=0.585 m (n=10, s.d=0.156 m) whereas G2 exhibited a higher threshold (0.687 m, n=8 s.d=0.0641 m). If expressed in terms of curvature ($C_{50} = 1/R_{50}$), the values for the two groups are respectively 1.817 m^{-1} (s.d.=0.466 m^{-1}) and 1.465 m^{-1} (s.d.=0.133 m^{-1}). These two values were significantly different from the initial curvature ($p < 0.002$ and $p < 0.000001$ respectively). The mean curvature thresholds for G1 and G2 were significantly different ($p < 0.05$): subjects from G2 reached a higher level of curvature discrimination, that is, they were able to sense the curvature direction of flatter surfaces. These results are visible on figure 4.5. This observation is corroborated by the total number of correct answers given by the subjects in each group: 18.3 s.d.=9.3 for G1 and 26.7 s.d.=6.1 for G2. An analysis of variance revealed a significant effect of the group on this value ($p < 0.05$, figure 4.5, center).

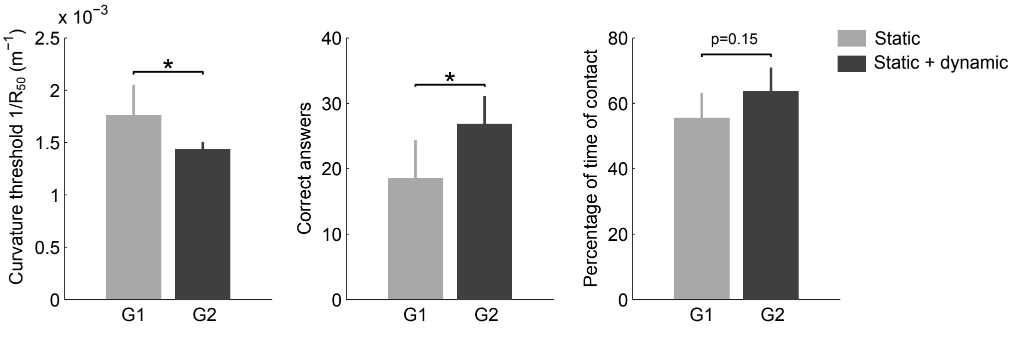


Figure 4.5 – Results of curvature thresholds (left), correct answers (center) and percentage of time in contact with the AVS (right) for both static and static + dynamic feedback groups. Significant differences are indicated with '*' ($p < 0.05$). Vertical bars indicate 95% confidence intervals.

Time of contact and response time

The time spent in contact with the plate was not significantly different between the two groups (see figure 4.5 right), although there seems to be a tendency towards a higher time for G2 (63.8% of the total exploration time) than for G1 (55.7%). As far as RT is concerned, there were no significant difference measurable, neither between the two groups, nor between responses given while exploring concave and convex curvatures. As expected, the analysis of RT did not bring major observation because of a large between-subjects variability. That might have been caused by the relatively long time window subjects had to give their answer, and the difficulty to restrain RT to 30 seconds by reminding this limit to them. Although, for some subjects, the RT show some timing and stimulus-dependent characteristics.

A first example is given in figure 4.6 for subject #2 (from G2) where the RT follows a learning-like evolution. The fitted model (black line) is an unconstrained nonlinear optimization of the three-parameter $a/(x+b)+c$ function of RT over trials. This curve shows that the subject established a strategy while learning the task, leading to a decrease of his RT, and this, in spite of the increasing difficulty along the experiment (increase of the curvature of the plate). This strategy also seems to be suitable for both signs of curvature.

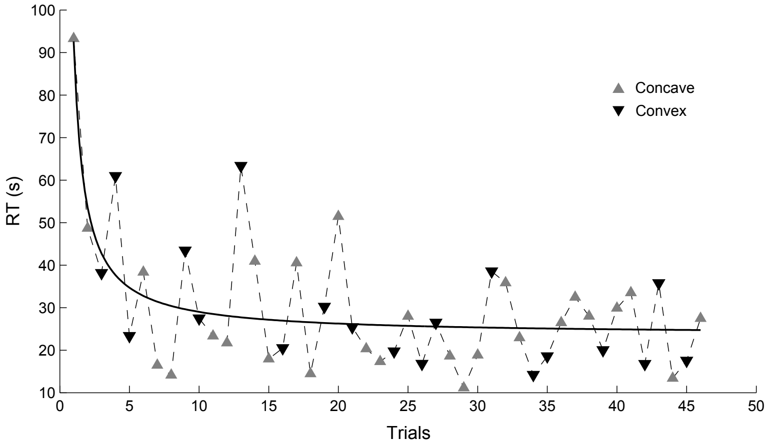


Figure 4.6 – Response times (RT) for subject #2 during the experiment, illustrating a learning behavior. Black and gray triangles stands for convex and concave stimulus respectively, the black line is an inverse function fitting of the data.

The second example, illustrated on figure 4.7, shows the case of subject #10 (G2 as well) who exhibits different RT ranges depending on the curvature direction. We observe larger RT for convex curvatures. As the difficulty increases with trials, this behavior is emphasized: the RT increase for convex curvatures (black triangles on the figure) whereas the concave RT (gray triangles) remain more stable. Any auditory-motor exploration strategy this subject did set up in the early trials, it turned out to be more effective for one curvature direction. This observation can justify the choice of a discrimination task as it is not obvious that the difficulty is the same in both curvature directions for the subjects.

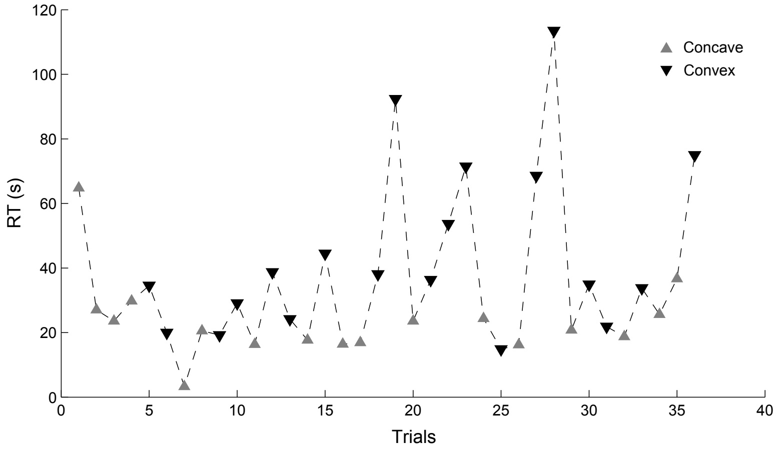


Figure 4.7 – Response times (RT) for subject #10 during the experiment, illustrating stimulus-dependent RT. Black and gray triangles stands for convex and concave stimulus respectively.

These two cases are a short snapshot of the diversity of behavior we could observe in this experiment. So far, the observations -more or less subject to between-subject variability- may be linked to underlying mechanisms driven by strategies that the subjects could have built while experiencing the AVS. In the next paragraphs we observe the individual gestures produced by the subjects as well as their personal feelings to examine these strategies of exploration.

Gestural strategies

By looking at the spatial trajectories performed by the subjects during the exploration of the plate, two gestural patterns appear clearly. On the one hand, some subjects performed vertical gestures, oscillating along the Y axis and ‘tapping’ or ‘bouncing’ on the plate. On the other hand, a second pattern exhibits lateral gestures, bended upwards on downwards, ‘sweeping’ the plate all along the X axis. We have to mention that these patterns present multiple variations across the different subjects and also some adjustments as the difficulty increases with time and learning occurs.

To illustrate the two patterns observed, some examples are shown figure 4.8. The first gesture, ‘tapping’, is shown on the left side (a and b), the second, ‘sweeping’, is shown on the right (c and d). The trajectories are all drawn with the same scaling; the red cross marks the center of the plate at $(x, y) = (0, H)$. One can easily notice the general aspects of the two gestural patterns and their main visual differences. On the left, the gestures remain mostly orthogonal to the surface, regularly and quickly repeated. They present a high number of vertical turnarounds, as if the subjects were ‘tapping’ or ‘groping around’ the surface. Subjects’ hands thus stay an important part of the time outside of the AVS. As the curvature gets smaller we observe an higher number of bounces on the plate. The second pattern, visible on the right side of figure 4.8, shows, oppositely, mostly lateral movements with less oscillations and larger segments in the AVS. The trajectories are kept horizontal (see c) or may be curved and parallel to the surface (d). Besides, the number of contacts with the AVS seems to be higher in the first pattern and the length (or time) of the motion in it seems lower.

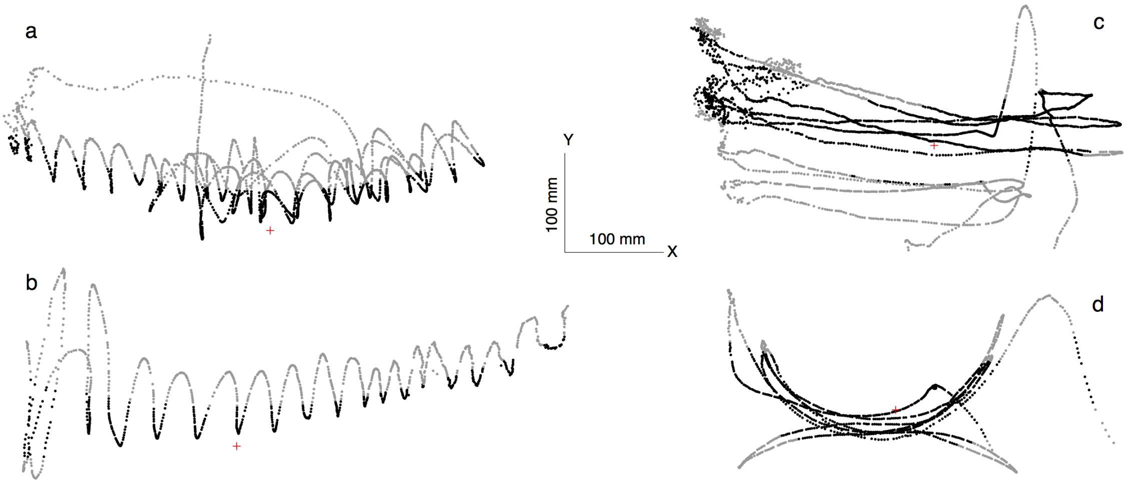
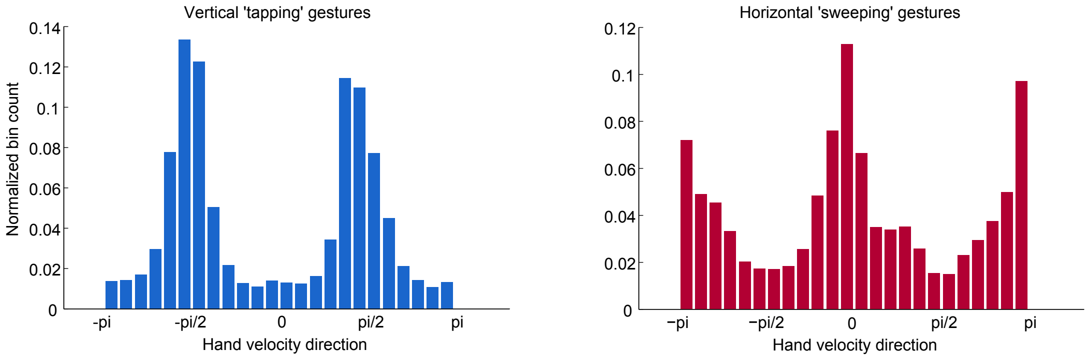


Figure 4.8 – Examples of spatial trajectories, showing the two different gestural strategies: a and b show 'tapping' movements on the surface; c and d show 'sweeping' gestures that are more lateral and curved. The red cross indicates the center of the plate. Each dot is a time sample (jitter not corrected), corresponding approximately to 10 ms sampling. The black dot are positions inside the AVS geometry. The plates are concave in the four examples.

Computing the direction of the velocity vector along the trajectory allows to examine the distribution of the gestures direction when being in contact with the plate. The direction of the velocity vector have been computed for every trial and the angle distribution is built between $[-\pi; +\pi]$, the 0° direction being aligned with the X axis of the setup (see figure 4.2) and respecting the trigonometrical conventions. Then the distributions are averaged across trials.

Figure 4.9 shows the typical direction distributions for the two strategies that can be identified, the vertical (tapping) and horizontal (sweeping) gestures. The distributions are normalized so that the sum of the bins count equals one. The left distribution exhibits a strong dominance of $-\pi/2$ and $+\pi/2$ directions (*i.e.* vertical), whereas the right one has minimum values at this angle and maximum for 0 and $\pm\pi$ (*i.e.* horizontal). This representation confirms the observations made on the spatial trajectories and clearly allows to discriminate the behaviors using movement kinematics.

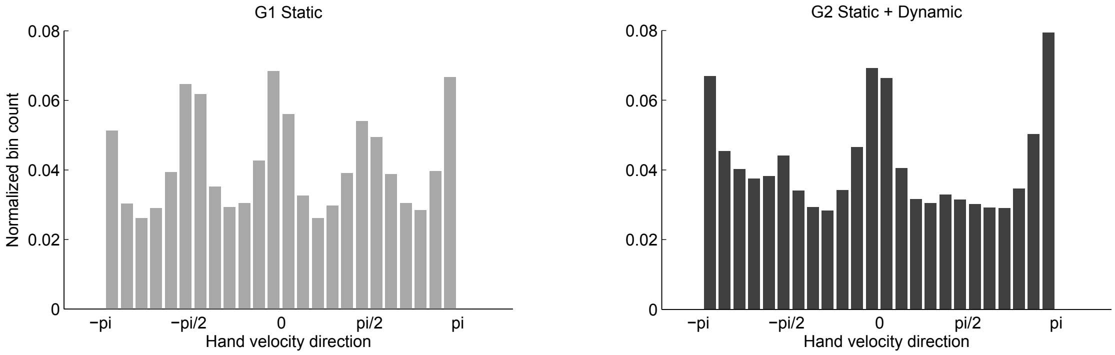
The respective average distributions for G1 and G2 were computed. The results are shown figure 4.10. A statistical analysis between the bin counts showed a significant effect of the angle on the distributions ($p < 10^{-6}$). The G1 group exhibits orthogonal velocities that the G2 does not have: the G1 group distribution has between 30 and 40% more $\pm\pi/2$ counts than G2, and they reach the floor level of the distribution. Horizontal velocities are nonetheless found in both cases. No statistical differences were found between the



(a) (b)

Figure 4.9 – Distributions of velocity directions for the two identified gestural strategies tapping (left) and sweeping (right). The relative weight of vertical and horizontal movements differs according to the strategy used.

two groups, due to important inter-subject variability, although the two signatures of the identified behaviors can be noticed in the distributions. This reveals that the two identified gestural strategies are not specific to one group or another. Furthermore, no relationship was found indicating a correlation between a gestural strategy and the discrimination threshold. Some subjects may also use both kinds of gestures to achieve the task, regardless of their performance level. As shown figure 4.10 there is though a tendency for subjects in G2 to produce less vertical movements while exploring the AVS.



(a) (b)

Figure 4.10 – Distributions of velocity directions; left is averaged for G1 subjects (static feedback), right is for G2 subjects (static+dynamic feedback).

Perceptual assessments

The questionnaire which was submitted to the subjects after they finished the experiment revealed that 80% of subjects in G1 felt some discomfort being deprived of sight to do the discrimination task. They were only 25% in G2. As far as their feelings are concerned, 50% of subjects in G1 reported they felt pseudo-tactile stimulations when interacting with the AVS. In contrast, 75% of subjects in G2 did report such sensations.

Among the diverse words the subjects employed to describe their feelings, we noted ‘tingling’, ‘pins and needles’, ‘vibrations’, ‘veil under the fingers’ or a ‘shape’ felt under the palm. It is interesting to notice that these words all refer to physical interactions and tangible experience. One subject even described that he felt *heat* on his finger pads when he was ‘touching’ the AVS.

4.6.5 Discussion

The experimental procedure we proposed here allowed to evidence that subjects were able to appraise the direction of curvature of an AVS using only auditory feedback, which is not trivial. In our case, the auditory feedback was not a supplementary feedback (or ‘augmented’ feedback) but precisely the core of the interaction where subjects could only rely on sound to move and explore the virtual object. This contrasts with other studies (Alonso-Arevalo et al., 2012) where sound was an additional signal for haptic exploration. The estimation of the curvature discrimination threshold we measured here with the palm position is only between two and three times larger than the threshold measured with a dynamic exploration of real objects with the finger, for example measured by Norman et al. (about 0.6 m^{-1} (Norman et al., 2013)). We then hypothesized that we could lower this threshold by creating an interaction with each fingers of the hand (as the device now offers a skeleton of the hand in the API), and making the quantity of information delivered proportional to the number of fingers in contact with the surface, as it would be the case with a tangible surface.

The fact that the threshold obtained by the group receiving both the position and velocity feedback is significantly lower than for the position-only group shows that providing this additional information helped the subjects to explore and sense the AVS. This result highlights that the velocity related feedback provides a valuable information for the interaction, in the absence of sight in that case. The continuous aspect of this feedback is also of importance as it is coherent with the idea of exploring a homogeneous portion of space. In contrast, the static feedback may lack of interaction with motion. This idea is confirmed by the perceptual feelings expressed by the subjects.

A link between the performance and the type of feedback has been shown, but the evidence that G2 subjects performed better because of a different exploration strategy is still missing, as some subjects using another strategy serve perfect counter-examples. The time of contact with the plate is not significantly different between the two groups. Nonetheless, we suppose that the lateral ‘sweeping’ type of exploration can increase this amount of time, as the higher average for the second group tend to indicate.

We clearly identified two exploration strategies of the AVS in terms of movement kinematics. This result shows that subjects were able to establish a deep link between their exploration - mainly proprioceptive - and the sound. The first strategy, ‘tapping’ the surface, consists in using the appearance of the sound as a positive trigger to memorize the hand position at this moment in order to build the 'map' of the AVS. This must be repeated a large number of time to cover the AVS length and obtain a mesh model fine enough to evaluate its curvature. As we were able to observe, the number of points and the length covered with the exploration should increase with this strategy as the curvature of the plate gets smaller.

On the other hand, the horizontal 'sweeping' strategy uses the sound disappearance as an indicator of exiting the plate. To find the curvature direction, the subjects performed several gestures parallel to the surface and measured the time they hear the sound before it disappeared. By performing concave and convex gestures, the subjects using this strategy simply had to choose the gesture which gave rise to the longer sonic interaction. These slightly curved gestures might have been favored by the tangential velocity sonification.

If the participants had been allowed to use vision, the differences between G1 and G2 would have probably been blurred as the position feedback would have been more easily associated with the position of the hand in space, making the additional feedback in G2 unnecessary. In addition, we hypothesize that the 'tapping' strategy would have been predominant in that case.

The concept we used seemed disturbing for some subjects when the instructions were given to them. They did not seem confident until they clearly understood the task, that is, after a few trials. Picturing a virtual surface, moreover reactive and sounding, without prior schematic visualization turned out to be a sensitive issue for the protocol. A familiarization procedure could be beneficial as the ten first trials may be needed just to understand the task. That being said, the majority of the participants recruited were involved and managed to perform the task.

The first outperforming subject (from G1) ruled out of the analysis reached a curvature threshold of 0.77 m^{-1}. This subject was able to describe precisely his strategy and it corresponded exactly to the 'tapping' one. To achieve this threshold, he monitored and remembered the flexion/extension of his shoulder in the sagittal plane, move after move at the superior face of the plate. The second subject (from G2) reached 0.62 m^{-1} and adopted the 'sweeping' strategy. He described that once he found the 'middle' of the AVS thickness, he produced horizontal movements in layers at various heights and searched for the longest auditory feedback he could perceive. He insisted that he used to this end the velocity feedback, and that the position feedback allowed him to concentrate on the idea of the virtual object. Both subjects thus achieved high precision by focusing on proprioceptive information guided by the sound.

These cases suggest that the experimental procedure should be adapted in the case of high scores. For instance, after a certain curvature threshold is passed, the thickness of the plate could be reduced so that the difficulty keeps increasing linearly. Preliminary tests should be carried out to assess the difficulty scale. The second issue is that these subjects contrast so much with the others that they had to be withdrawn from the statistics. If a pre-learning phase is adopted, it could be used to rule out the subjects who do not get used to the task. Finally, these results were obtained with a small number of subjects and will need to be confirmed with a larger panel, and testing the potential influence of specific trainings (video-games, sports, etc.) to complete this proof-of-concept study.

4.7 Conclusion and perspectives

The experiment we presented here allowed us to develop the concept of Auditory Virtual Surface and apply it to investigate the substitution of touch by sound. The results showed it is possible to make subject perceive the curvature of a virtual object using only sound. A feeling of ‘touch’ was even felt by the subjects thanks to this auditory-proprioceptive/tactile coupling, even though we did not use a physical sound model of the interaction, which could have been thought as more ‘realistic’. We also showed that a dynamic sonic interaction can enhance the perceived modality and the induced feelings. This reveals the crucial role of the perception-action loop in a gesture/sound interaction in the case of sensory substitution.

Upfront and down-front of the sound triggering were used differently between the two strategies that were observable. One unveils the AVS geometry from the outside of it as a hollow mold, by building a discrete model, and one explores it from within (as the negative of the mold) in a more continuous way. This highlights how diverse and rich the personal interpretation of a interactive sonification can be. The concept and the experimental protocol presented illustrate a new approach of the relationship between proprioception and sound, and offer a framework for understanding how the auditory-proprioceptive and tactile couplings could be integrated for spatial exploration.

4.7.1 Related questions

The paradigm of sensory substitution could benefit from movement or object sonification and auditory-motor interaction, at both fundamental and applied levels. Sensory substitution devices that use the auditory channel often attempt to render static or dynamic scenes with few interaction (Bach-y Rita and W. Kercel, 2003). We think that it is worth adding an interactive (movement-based) layer to this paradigm. The field of virtual reality relies partly on this notion of interactivity to access the realism needed (the reader can refer to the extensive literature of this domain), especially when applied to rehabilitation either physical (Cameirao et al., 2007; Correa et al., 2007; Holden, 2005; Lehrer et al., 2011a; Lehrer et al., 2011b; Rose et al., 2005), or mental (Timmermans et al., 2004). Merians et al. (Merians et al., 2006) underlines the need for home-based, thus, low-cost and practical, virtual reality environments for physical rehabilitation and sensorimotor training. We demonstrated with this study that a low-cost, interactive auditory scenario could settle sensorimotor learning and induce gestural patterns and sensory substitution effects with possible synaesthesia.

The vocabulary used to describe the gestures performed in the two strategies observed can be seen as an analogy with the vocabulary of the (electric) guitar playing techniques. The 'tapping' technique is a playing mode where the notes are produced by fretting the strings with both hands (often called 'hammering' when it is performed only with one hand). An instrument is actually based on that playing technique, called the Chapman stick[3], which allows for specific sounds, particularly concerning the attack of the note. The 'sweeping' technique allows to play several strings quickly one after the other with a downward or upward global motion without changing the attack side of the pick. This technique is optimized for a fast playing with a smoother sound. It requires though a strong coordination with the other hand on the fretboard. The gestural patterns identified here could be unchanged if one would explore a real and tangible object. The fact that the action-perception coupling drives such gestures opens new pathways to explore virtual auditory spaces and interactions, including new designs for digital musical interfaces (as tactile information plays a role in the temporal accuracy of musical playing (Palmer et al., 2009)).

4.7.2 The virtual water bowl

We showed with the previous study how moving and listening could build an auditory image of the interaction, even in the case of virtual reality. Similar setup and protocol could be used to observe behavior in the case where the AVS is a physically modeled object or body, whose sonic outcomes are only perceivable. The reality is again distorted but imaged with the sonic interaction. By focusing the attention on the auditory consequences of the action, and without sight, the user may again modify his behavior and develop a different motor strategy. An example of setup has been designed to illustrate this point with water.

The idea came when our team was approached by a foley artist to study the feasibility of using a motion capture interface to design and control foley sounds. He specifically complained that some family of sound effects or textures were extremely difficult and unpractical to manipulate. For instance, he mentioned the water sounds that need to be recorded in specific facilities (for practical reasons we easily understand). Although one of the skills of a foley artist is precisely to bypass this problem (they often create the sound of a flame by waving a heavy piece of fabric very close to the microphone, for example), the artist insisted on that point. We developed a prototype allowing to interact with a virtual water bowl in the air, using a LeapMotion™ sensor. The user can splash the water surface and produce sweeping movement in the water, beneath the surface.

Two sound engines were used, one for splash sounds and one for underwater sounds,

---

[3] `http://www.stick.com/`

both developed with corpus-based concatenative synthesis (Schwarz and Schnell, 2010; Schwarz, 2012). This technique consists in computing audio descriptors (like duration, loudness, energy, spectral centroid, etc.) on pre-recorded sound samples and playing sound segments best matching given audio characteristics. The sound selection can be achieved with decision trees, allowing to tune weights and probability between selected sounds and data input. Our example used the *knn* (k-NN unit selection based on a kD-tree algorithm) and the concatenative synthesis modules from the Mubu series of Max/MSP objects (Schnell et al., 2009). Gestures were captured sensing the 3D position and velocity of the hand above the device (where the water should be). Sound materials (original recorded splash and water sound) came from the French ANR Topophonie project[4].
For splash sounds, we computed the energy and the velocity direction of the hand when 'touching' the surface of water. These parameters were used for the selection of sound samples, based mainly on the length of the splash sound and (to a lesser extent) the energy of the sample (each parameter weight can be adjusted). Finding the appropriate audio descriptors distribution to select the sounds was empirical design. The duration of a splash sound is fairly correlated with the energy of the gesture which produced this sound; as splashing into (real) water is perturbing a physical system until it reaches equilibrium again, the more energetic the gesture, the longer the sound produced. Transition probabilities were also adjusted in order to avoid repetition of the same sound sample and ensure a minimum variability (as the sound corpus may not be infinite).
For underwater sound, a similar approach was chosen, adding special effects on water sounds to imitate underwater recordings. Peak velocity of lateral sweeping movements underwater was used to trigger the synthesis. A double switch audio gate allowed to change smoothly between the two engines when the hand crossed the surface (in order not to disrupt the sound texture). Finally a stereo panning mixed the output from left to right according to the hand position in the interactive zone.

This very playful prototype allowed for an embodied interaction with a virtual volume of water in both surface and underwater configurations (as if the ears of the user were actually in the bath tub). The interested reader can see a demonstration video at this address[5]. This setup can be used, for instance, by mixing the plate curvature protocol presented in this chapter with physical sounds (*e.g.* water), to observe the gestural adaptation occurring while exploring such a virtual sonic object. We believe that sound texture audio engines are of great interest for developing engaging interactive scenarios. This prototype shows that powerful available audio technology, controlled with a commercial motion capture device can be a practical technical solution to develop portable and low-cost rehabilitation or sensory substitution interactive scenarios.

---

[4] http://topophonie.com

[5] `https://youtu.be/mTlizSOJbso`

Chapter 5

Continuous sonification in a two-dimensional visuo-manual tracking task

E. O. Boyer[1,2], F. Bevilacqua[1], P. Susini[1], S. Hanneton[2]
[1] IRCAM, STMS-CNRS-UPMC, Paris, France
[2] LPP, Paris Descartes CNRS UMR 8242, Paris, France

Abstract: The use of real-time continuous auditory feedback for motor control and learning is still under-studied and deserves more attention regarding its potential applications. This paper presents the results of three experiments studying the contribution of sonification to a visuo-manual tracking task and assessing its benefits on motor learning. Results show first that auditory feedback is beneficial for performance, whether it is error-related or task-related. Second, when the feedback modality reflects the user's movement only, the positive effect remains after feedback is removed and performance continues to improve if feedback presence alternates. Furthermore, the presence of auditory feedback increased the average level of energy in the participants' movements. A retention test showed that sonification of the user's motion helped consolidating learning. These results confirm that a continuous auditory-feedback can be beneficial for movement training and that the effect of a user-related feedback does not only refer to the error but can have a deeper effect on the auditory-motor loop. The sonification of the user's own dynamics thus appears as a very promising path for developing engaging interactive scenarios.

Keywords: Tracking movements, Auditory feedback, Continuous sonification, Learning.

Article submitted to Learning and Memory.

5.1 Introduction

Continuous visuo-manual tracking tasks have been widely used in neuroscience research as an experimental paradigm to investigate human motor control and behavior (Craik, 1947; McRuer, 1980; Miall et al., 1993). During a tracking task, the participant has to pursue continuously a moving target that can follow a periodic (predictable) or noisy (not predictable) trajectory. This continuous task allows for well-controlled experiments with a limited number of parameters since generally only two sensory modalities are involved: vision and proprioception. Moreover, the movement regulation loop can be simply formalized by considering a single output command (the displacement of the hand on the interface) that responds and anticipates to few sensory inputs (for instance the distance between the target and the pointer position). Studies where authors tried to model the behavior of participants brought an essential outcome: the tracking behavior can not be modeled by a linear input/output function (Hanneton et al., 1997). In other words, the movement of the participants contains frequencies that are not present in the target trajectory. Even if this task may be considered as simple, there are up to now no satisfying model of the participant's behavior. The purpose of the present study is to investigate more particularly the perception/action coupling in a visuo-manual tracking task with various types of auditory feedback. Specifically, the effects of a real-time continuous auditory feedback on performance and learning are assessed.

The use of augmented feedback (Hartveld and Hegarty, 1996) to enhance motor control has been largely studied over the past decades. Nevertheless, few studies have considered the auditory modality as augmented feedback. For example, Effenberg (Effenberg, 2004) highlighted the potential of sound feedback to enhance movement perception. Kagerer and Contreras-Vidal describes particular phenomenons using auditory-motor transformations (Kagerer and Contreras-Vidal, 2009). They observed that a new internal model built after exposition to a visuo-motor perturbation can be used by the auditory-motor system immediately after exposure. Sigrist et al. (Sigrist et al., 2013) presented a review of augmented feedback techniques for motor learning and pointed out the high potential of auditory feedback. Technical difficulties and choices in designing sound feedback seem to explain the relative simplicity and narrow range of published experiments and applications. Nevertheless, promising results showed that sound feedback, among others, can help learning movements (Sigrist et al., 2014; Boyer et al., 2014) or to control the orientation of an object (Rath and Schleicher, 2008).

Providing extrinsic feedback while learning a motor task can lead to a potential dependency on the feedback (van Vliet and Wulf, 2006). This observation is sometimes referred to as the guidance hypothesis (Salmoni et al., 1984). It has been suggested to reduce the rate or time of feedback presentation to minimize this effect (Winstein and Schmidt, 1990; van Vliet and Wulf, 2006). But feedback dependency seems to be linked to the task and the context: guidance can develop during learning if the context is favorable to a spatial or visual representation of the action (Buchanan and Wang, 2012), suggesting training with feedback always present is not necessarily detrimental. Very few studies examined this effect with auditory feedback though. Ronsse et al. (Ronsse et al., 2011) showed that sound feedback can be used to learn a bimanual coordination pattern, and that auditory feedback did not lead to dependency on the feedback at the end of practice, unlike the use of a visual feedback (Avanzini et al., 2009). In the context of robot-assisted rehabilitation, several studies showed that sound feedback is efficient to complement visual feedback. Rosati et al. (Rosati et al., 2012) reported studies on unidimensional tracking with a joystick, comparing the sonification of the target velocity and the tracking error. They found that auditory feedback based on task parameters improved performance during an unpredictable task. Their results also showed that sonification of the tracking error did not have positive effect on performance and tend to deteriorate adaptation to a visual perturbation. This study did not include a control group and also focused on input interface mapping.

Similarly to Rosati's study, we report here the comparison between three different sonification strategies in a visuo-manual tracking task. The three sonification strategies are based on the same sound synthesis system and exhibit identical acoustic features. As explained below, the difference resides only in the sonified quantity. The first sonification is related to the instantaneous distance between the target and the pointer. It is *per se* an augmentation of the available visual feedback and will be called 'error-related' feedback (*Error*). The second and third sonification strategies are related to instantaneous velocity signals: the 'task-related' sonification (*Target*) reflects the velocity of the target, and the 'user-related' the velocity of the manipulated pointer (*Pointer*). We emphasize that the *Pointer* sonification provides information only about the participants' movement, regardless of the task. To our knowledge, it is the first time that the effect on motor control of a dynamic, continuous and user-related feedback is tested in a tracking task.

The following hypothesis are tested in this study: a- real-time and continuous sonification, as an augmented feedback during a visuo-manual tracking task, can improve the precision of tracking, b- the parameter used for sonification (error, task, user, etc.) can cause different effects on the tracking performance, c- feedback dependency effects may occur, and learning stability may vary according to the type of sonification, and d- the augmentation of the action-perception loop with sonification can modify hand movement features in the context of the task. Three experiments were carried out with the help of 108 participants (see Materials and Methods). The effect of the three sonifications were tested and compared in experiments 1 and 2. Experiment 2 allowed to test feedback dependency. The third experiment focused on the two velocity-related sonifications. It included a reduced exposition to the feedback, and a 24-hours retention test to observe learning stability.

5.2 Materials and Methods

5.2.1 Subjects

A total of one hundred and eight participants volunteered for the study (36 participants for each experiment). They were aged from 18 to 70, with an average of 29.2 ± 9.9 years old. The gender balance was 40.7% female 59.3% male. Table 5.1 presents the detailed statistics about the participants in the different experiments. Every participants participated in only one experiment and in one group exclusively. Exclusions criteria were: diagnosed hearing loss, physical impairment of the dominant arm and hand, color-blind condition or inability to distinguish the colored dots on the screen. Subjects were asked to rate their video games and sports practices on a custom five-point Likert scale ranging from 1-never and 5-every day. All participants were healthy and had normal hearing. This study was carried out in accordance with the Declaration of Helsinki and approved by the health research projects ethics committee of Paris Descartes University (International Review Board number 20142700001072). All participants gave written informed consent. Participants in Experiment 3, which last over two days, received a compensation.

Table 5.1 – Age, gender, self-reported video games and sports practice for all the participants

	N	Age	Gender	Video games practice	Sports practice
Experiment 1	36	30.8 ± 10.3	30.6% F	2.3 ± 1.3	2.6 ± 1.0
Experiment 2	3x12	27.4 ± 9.6	44.4% F	1.9 ± 1.1	3.1 ± 1.2
Experiment 3	36	29.3 ± 9.7	47.2% F	1.9 ± 1.1	2.3 ± 1.1
Total	108	29.2 ± 9.9	40.7% F	2.0 ± 1.2	2.6 ± 1.1

5.2.2 Experimental setup

The three experiments shared the same setup. Auditory feedback was delivered through headphones (AKG K271 MKII) that people wore in all the conditions. Experiments 1 and 2 were carried out in quiet offices at Ircam-Centre Pompidou and in the STAPS department of Paris Descartes University, France. Experiment 3 was carried out in a double-walled sound-insulated booth at Ircam-Centre Pompidou.

In all the experiments, participants were sited in front of a desk with a graphic tablet on top (Wacom Intuos2, 304 x 228 mm, with XP-501E stylus). The visual environment was displayed on a Samsung SyncMaster 2053BW screen driven by an ATI Radeon HD2600XT 256 Mo graphic board. The computer used was a Mac Pro 2x2.8 GHz and 6 GB RAM running OSX 10.8.5. A program built under the Max environment (Cycling'74) allowed for the experiment control, real-time data processing and recording, as well as visual display and auditory feedback production. As the participants moved the stylus on the tablet surface, the (x, y) position data from the tablet translating the position of the cursor on the screen were analyzed and recorded at a sample rate of 100 Hz. The graphics were rendered using the *jit.jl* Max objects collection.

The overall latency of the system between the stylus moving on the tablet and the sound feedback generation was assessed by recording the contact sound of the stylus on the tablet surface with a microphone and collecting the time delays of the subsequent events in the processing chain. The total measure was 31 ms taking into account the audio driver latency.

Visual display

The target and pointer visuals were represented on the screen by respectively red and green 10.8 mm diameter plain circles on a black background. They were rendered at 60 Hz on 1680x1050 pixels (433.4 x 270.9 mm), the maximum resolution of the screen. The position of the target was generated with random numbers at 100 Hz and low-pass filtered to get a reasonably smooth target point, yet difficult enough to follow. Filters used were second-order recursive linear filters with 0.6 Hz cutoff frequency, unit normalized gain and q-factor. Three cascade filters were used to get a -36 dB/oct slope.

The program allows for either the generation of random trajectories (computed online) or for playing back pre-recorded trajectories. For Experiment 3, three different trajectories have been pre-recorded corresponding to blocs 1-2-3, and played back in the sessions.

Auditory feedback

The auditory feedback was generated in real-time using white noise filtered with a resonant filter (Max object *reson~*) with a resonance factor of $Q = 23$, its center frequency varying between between 80 and 4000 Hz. This range is mapped to the minimum and maximum range of the varying parameter of each feedback type described below.

Three types of feedback were designed for the experiment sharing that same architecture. For the first one, the error-related feedback (*Error*), the sound is modulated by the Euclidean distance between the target and the participants' pointer on the screen. The two others use the velocity of the target (*Target*) and the pointer (*Pointer*) to control the filter frequency. The maximum frequency of the filter command was reached at 120 cm.s^{-1} on the tablet.

Data analysis

The recorded trajectories produced by the participants have been low-pass filtered at 8 Hz with a Gaussian filter before further signal processing analysis. The RMS tracking error

(RMSE) was computed as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i}^{N} err(i)^2} \tag{5.1}$$

where $err(i)$ being the instantaneous Euclidean distance between the coordinates of the target (x_t, y_t) and the pointer (x_p, y_p):

$$err(i) = \sqrt{[x_p(i) - x_t(i)]^2 + [y_p(i) - y_t(i)]^2} \tag{5.2}$$

The normalized energy in the movement for a block was defined as the amount of energy in the tangential velocity signal normalized by the same quantity for the target:

$$E = \frac{\sum(v_p(t))^2}{\sum(v_t(t))^2} \tag{5.3}$$

where $v_p(t)$ and $v_t(t)$ are the tangential velocity signal for the pointer and the target respectively. E thus gives an image of the supplementary energy in the participants motion compared to what is theoretically needed to follow the target.

5.2.3 Experimental procedure

After reading the general instructions and giving their informed consent, participants chose their dominant hand for the task. The participants performed tasks for a total time of 12 minutes, by blocks of 3 or 4 minutes with a 1 minute break between each block, depending on the experiments. The following sections describe the different experiments that combined different feedback conditions:

- *NoAudio*: no auditory feedback
- *Error*: error-related sonification
- *Target*: target velocity sonification
- *Pointer*: pointer velocity sonification

Experiment 1 tested the three auditory feedback in different orders versus the *NoAudio* condition for 36 participants. In experiment 2, three groups of participants (one for each feedback) alternately received feedback and the *NoAudio* condition. Finally, in experiment 3 only *Target* et *Pointer* conditions were used to train two groups of participants with 50% feedback presence and a control group who was under the *NoAudio* condition. This experiment included a 24 hours retention test for the three groups, carried out on the next day without auditory feedback. At the end of each experiment, participants were asked to rate their perception of the task difficulty, the auditory and visual fatigue. Difficulty and fatigue are rated on a five-point Likert scale (for experiment 1 and 2 only). They were also invited to give free comments about their feeling.

5.3 Results

5.3.1 Experiment 1

In experiment 1, the participants were asked to perform the task during four blocks of 3 minutes. The first block was the *NoAudio* condition. Then, each participant was exposed to three different blocks corresponding to the three feedback conditions: *Error*, *Target*, and *Pointer*. In order to test any possible order effect, each participant was assigned to one of the 6 different possible groups: E-T-P, E-P-T, P-E-T, P-T-E, T-E-P, T-P-E. An analysis of variance on the RMSE tracking error with the group factor revealed no effect of the order of the sonification conditions. This allowed us to average the tracking error results across groups to study the effects of the different sonification conditions. Statistical analysis showed a significant effect of sound on the tracking error ($F(3, 90) = 13.563, p < 0.005$) between block 1 *NoAudio* and the subsequent sonification blocks. The presence of auditory feedback led to a substantial decrease of the tracking error ($p < 0.00001$, Bonferroni post-hoc tests), by 12.8%, 13.7% and 12.3% for *Error*, *Target* and *Pointer* conditions respectively. These results are shown on the left side of figure 5.1. Overall, no significant difference was found between the three types of sonification regarding the general performance.

The amount of normalized energy in the movement E is superior to 1 for all conditions (see figure 5.1), indicating as expected that during the task, the closed-loop regulation makes the participants overshooting the target position. An example of motion data is shown figure 5.2: the pointer often overtakes and crosses the target position, and ‘catch-up’ saccades are even visible, like at 7 seconds on the figure. The speed signal visibly contains more energy and shows the way participants try to correct the position error. The three sonification conditions had a significant effect on E ($F(3, 90) = 15.110, p < 0.00001$, right side of figure 5.1) and Bonferroni post-hoc tests confirmed that these conditions exhibited higher E values ($p < 0.0001$) than the *NoAudio* condition. Subjects movements were thus more energetic with sonification, nevertheless, no difference was found between the three sonification conditions.

Subjects reported no particular auditory fatigue while hearing the auditory feedback (1.1 ± 0.4 over 5) and experienced average visual fatigue (2.4 ± 1.2). A small number of participants, who seemed more strained, reported hand fatigue or contraction by the end of the experiment and took advantage of the breaks between the blocks to stretch and relax their hand. Subjects generally rated the task easier to perform when sound feedback was present: 2.6 ± 0.9 with sound against 3.6 ± 1.0 without. No significant difference in the difficulty ratings were found between the feedback conditions. Ten people spontaneously declared by the end of the session that sound helped them to perform the task.

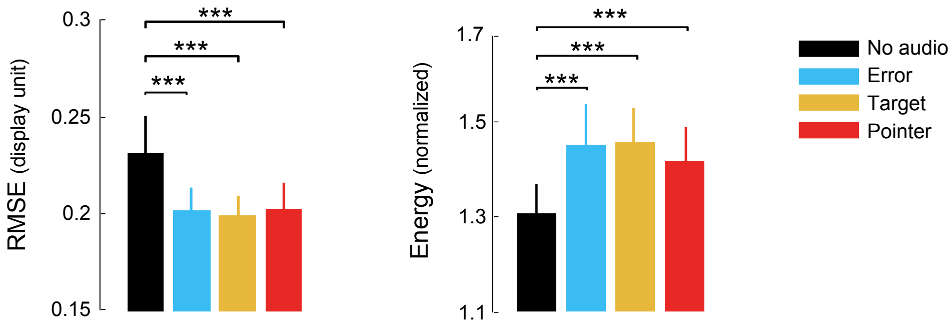


Figure 5.1 – Tracking error RMSE and movement energy E for each condition in experiment 1. The error bars indicate 95% confidence intervals. Pairwise comparisons and significance levels are indicated above the bars, $*** = p < 10^{-3}$.

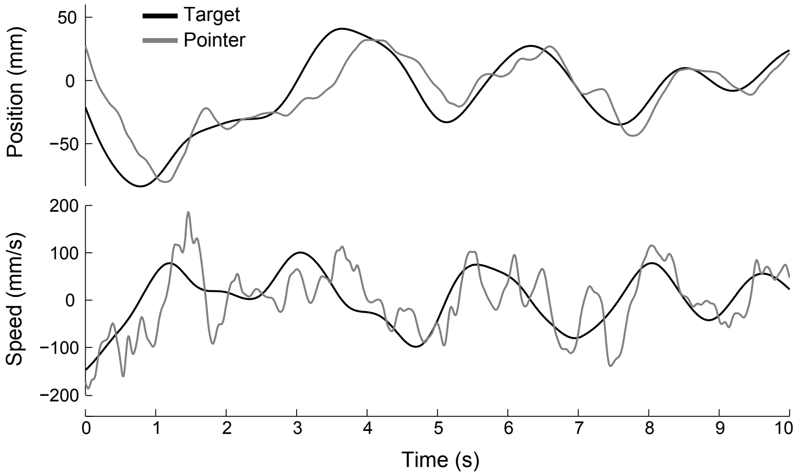


Figure 5.2 – Typical example of tracking motion (top) and velocity (bottom) along one dimension. The pointer (gray line) crosses the target trajectory (black) showing error control, also visible in the highly energetic velocity.

5.3.2 Experiment 2

In experiment 2, participants performed the task during 4 blocks of 3 minutes, each block alternating the *NoAudio* condition with one of the sonification conditions: *Error* or *Target* or *Pointer* (12 participants for each). Contrary to the first experiment where effects of practice and order cannot be discriminated, the auditory feedback were tested separately in this experiment. The participants were divided in two groups: the first group started with the *NoAudio* condition and the second one started with one of the sonification conditions. In each group, participants alternated twice between a sonification condition and the *NoAudio* condition, which we call the repetition factor.

In none of the series the order of presentation had an effect on the RMSE tracking error (effect of the group factor). There were no interaction between the group factor and sonification. Data were thus analyzed merging the two groups for each series. The statistical analysis included the measure of the sonification effect and repetition effect on pair of blocks as each condition was presented twice to the participants. In this analysis, the sound had a significant effect on performance for the *Error* and *Target* conditions ($F(1, 11) = 15.185, p < 0.005$, and $F(1, 11) = 25.051, p < 0.005$ respectively; as shown in Figure 5.3). Nevertheless, the repetition factor had an effect only for the *Pointer* condition ($F(1, 11) = 16.085, p < 0.005$). This observation is confirmed when performing the analysis on the 4 blocks (*i.e.* taking time progression into account). The three sonification conditions were found to have a significant effect on the RMSE tracking error compared to the *NoAudio* condition: $F(3, 30) = 5.3752, p < 0.005$ for *Error*, $F(3, 30) = 4.3347, p < 0.05$ for *Target* and $F(3, 30) = 6.6432, p < 0.005$ for *Pointer*. In the case of the *Pointer* sonification, we note that the error was continuously reduced block after block. Significance obtained by pairwise comparison results with Bonferroni tests are indicated above the bars in figure 5.3. The improvement with *Pointer* feedback is slower but seems more persistent and more robust to feedback removal than with the other types of sonification.

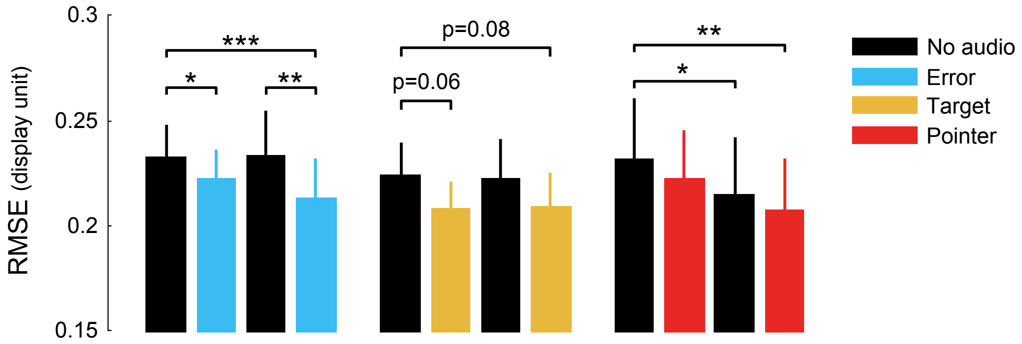


Figure 5.3 – Tracking error RMSE for each block and condition in experiment 2. he error bars indicate 95% confidence intervals. Pairwise comparisons and significance levels are indicated above the bars, *: $p < 0.05$, **: $p < 10^{-2}$, ***: $p < 10^{-3}$.

The presence of the auditory feedback had a significant effect on the normalized energy E ($F(3, 99) = 4.7917, p = 0.005$). The energy was increased with sonification, +7.3% on average. When analyzing feedback separately, the sonification affected E significantly for *Error* feedback ($F(1, 11) = 7.7221, p < 0.05$), significant thresholds were not reached for *Target* and *Error*, although a general increase with sonification is visible as well. *Pointer* sonification led to slightly higher E, 1.33, against 1.25 and 1.28 for *Error* and *Target* respectively, but without significance. The analysis when coupling sound and repetition factors led to the same observations and significance levels.

Concerning self-reports on perceived difficulty, no significant effect of sound nor repetition was found for *Error* and *Target* sonifications. For the *Pointer* group on the other hand, the difficulty was rated lower in the sonification conditions (2.5 versus 3.0, $F(1, 11) =$

$6.6000, p < 0.05$). The difficulty also decreased significantly during the experiment for this sonification ($F(3, 33) = 3.7692, p < 0.05$), from 3.3 to 2.5 ($p < 0.05$, Bonferroni post-hoc test). Subjects generally reported that they felt more "comfortable" or "focused" with the sound but admitted being unable to tell if their performance actually improved. Auditory and visual fatigue were rated similarly to Experiment 1 (2.6 ± 1.0 and 1.1 ± 0.5 respectively). The observations concerning hand fatigue are valid in this experiment too.

5.3.3 Experiment 3

Experiment 3 focused on the *Target* and *Pointer* sonification conditions, which are the two auditory feedback conditions based on velocity data. Subjects performed the task during two sessions, a *Training* session on day 1, and a 24-hours retention test (*Post* session) on day 2. Both sessions contained 3 blocks of 4 minutes. Contrary to the previous experiments, participants received auditory feedback 50% of the time in each block during the training session: the sonification conditions alternated with the *NoAudio* condition every minute *within* a block. On day 2, participants were asked to perform the same task but without auditory feedback. The participants were separated in three different groups corresponding to 1) *Target* condition 2) *Pointer* condition and 3) a control group who never received auditory feedback.

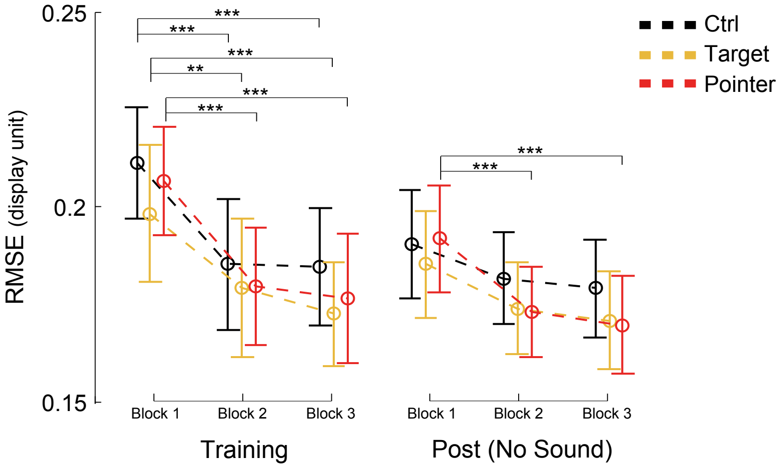


Figure 5.4 – Tracking errors RMSE for each block in Experiment 3 during Training and Post sessions. Error bars indicate 95% confidence intervals. Main pairwise comparisons and significance levels are indicated above the lines, **: $p < 10^{-2}$, ***: $p < 10^{-3}$. In the Training session, error is significantly reduced for the second and third blocks. In the Post session, only the *Pointer* sonification group improved performance during the retention test.

The block factor (3 levels) and the session factor (*Training* and *Post*) both had a significant effect on the tracking error ($F(2, 66) = 67.314, p < 0.0001$ and $F(2, 66) = 7.8895, p < 0.01$, respectively). The first block of each session was always the less successful ($p < 0.0001$ for both sessions). Figure 5.4 shows the RMSE results and significant post-hoc tests; changes in the tracking error shows a learning effect during both sessions.

Significant RMSE decreases for blocks 2 and 3 are observable in the *Training* session for the three groups ($p < 0.01$). The three groups exhibited a continuous error reduction of in this session: -13.0% for the *Target* group, -14.5% for *Pointer* and -12.5% for the control group, showing a general learning effect. During this session, no significant difference was found between the three groups. During the second day, a consolidation effect of the performance is observable: the error is globally lower than on day 1 ($p < 0.05$) but so is the amount of learning: -13.3% error reduction for *Training* and -8.4% for *Post* session. The post-hoc tests on day 2 revealed that learning is significant only for the *Pointer* sonification group ($p < 0.0005$ -11.5%, -7.8% for *Target* and -5.9% for the control group). The performance improvements between consecutive blocks or sessions are very sensitive to individual and within group variability. It must be emphasized that these results are obtained without ruling out any participant.

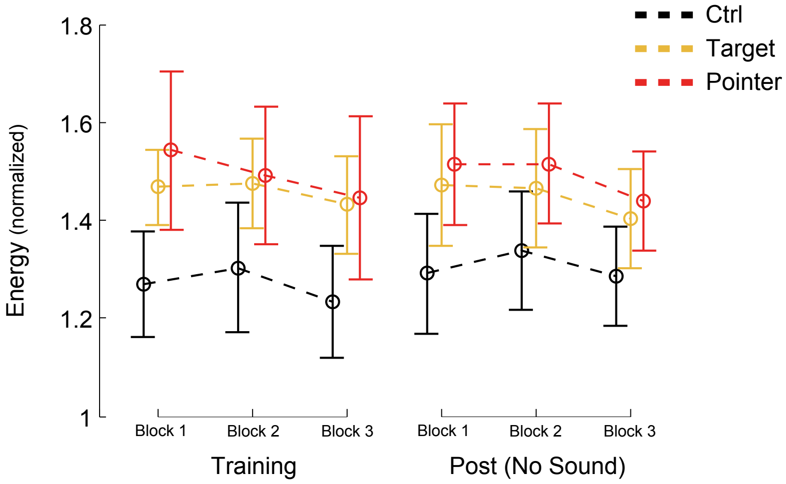


Figure 5.5 – Movement energy E for each block in Experiment 3. Error bars indicate 95% confidence intervals. The Control group exhibits significantly less energy than the two velocity feedback groups ($p < 0.05$).

Figure 5.5 shows the energy values and confidence intervals for experiment 3. Statistical analysis revealed a significant effect of block on the energy ($F(2, 66) = 8.8023, p < 0.005$), the last block of each session appearing less energetic but in a non significant magnitude. The group factor had a significant effect on the energy ($F(2, 33) = 3.3966, p < 0.05$). The auditory feedback groups exhibited significantly more energy than the control group: +12.8% for *Target* and +15.9% for *Pointer* on average, $p < 0.05$ Bonferroni post-hoc tests. However, the differences between these two feedback groups were not significant. No statistical difference was found in the energy between the two sessions.

A linear correlation between the tracking error RMSE and the movement energy E was computed for each of the 6 blocks of experiment 3. The normalized correlation coefficients are shown in figure 5.6 for the *Pointer* and control groups (the correlation was not significant for *Target* group). The p values of the correlation are indicated with '*' for $p < 0.05$ across the blocks. The correlation between the error and the energy in the movement for *Target* feedback increases along the experiment and evolves much faster than for the control group, especially during the *post* session. The correlation does not increase much during the experiment for the control group and is significant only in the first block of Post session.

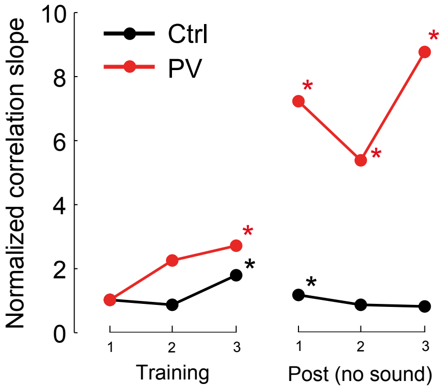


Figure 5.6 – Evolution of the correlation slope between RMSE and E for each block in experiment 3. The values have been normalized by the magnitude in the first block to observe the evolution with time. The '*' indicates $p < 0.05$ for the correlation.

5.4 Discussion

Performance and learning

The results of experiments 1 and 2 taken together show that auditory feedback helped improving performances in a visuo-manual 2D tracking task, which confirms our first hypothesis. Experiment 2 showed that *Error* and *Target* sonifications led to similar performance results, including concerning persistence of performance when feedback is turned off. Both groups exhibited a performance drop during blocks without auditory feedback, either they started the series with or without the feedback. However, the *Pointer* sonification group did not proved sensitive to feedback absence between consecutive blocks. Interestingly, this sonification is independent of both the target state and the performance. This result addresses our second and third hypothesis; it suggests that in this context, auditory feedback not directly related to the task is nonetheless integrated and can improve performance. Moreover, sonification of the user's movement helped the participants develop a learning process robust to changes in feedback conditions. *Target* and *Pointer* sonification led to similar performance improvements when auditory feedback was available only 50% of the time in each block. The retention test (preformed without auditory feedback) showed that the group trained with *Pointer* sonification increased significantly his performance as well, 24 hours after training, which was not the case for the other groups. This result can be connected with the robustness of the performance observed in experiment 2.

Experiments 2 and 3 allowed to observe feedback dependency (guidance hypothesis (Buchanan and Wang, 2012; Ronsse et al., 2011)), either directly on performance or during the retention test. It appears that error-related and target-related sonifications if always present while training, lead to a feedback dependency and to a weaker retention if available 50% of the time within blocks. The pointer sonification, on the contrary, led to continuous improvement. The stronger learning curve observed during the retention test (compared to *Target* and control groups) shows that training with this sonification allowed for performance retention in a longer term. The relatively slow improvement may be due to the task itself: at first sight the task seems simple, but improving performance during four minutes-long blocks proved to be demanding.

The fact that task-related as well as error-related auditory feedback significantly improved performance differs from the observations by Rosati et al. (Rosati et al., 2012), where an error sonification did not prove to help for the tracking task. Interestingly the authors also noticed, in the case of a visuo-motor perturbation that participants relied more on the sound from the visual target space than from the controller space. The visuo-motor perturbation introduced may have blurred the link between movement and the subsequent feedback. Compared to our results, this can confirm the necessity for the user-related feedback (*Pointer*) to be ecologically mapped to the user's movement and stable in time, in order not to disrupt the auditory-motor coupling that the user is developing.

In the particular case of visuo-manual tracking, vision provides instantaneous error evaluation and, within a larger time window, target velocity. As it focuses much of the participant's attention, auditory feedback should bring valuable additional information for the sensorimotor system to use it. We suppose that the *Pointer* sonification provides direct information on the arm dynamics, that is more easily integrated than the vision of the pointer's dot on the screen. It might be the parallel processing of visual (position of the target) and auditory information (imaging the user's movement) that allows for a more robust implicit learning in the case of the *Pointer* sonification.

Motion energy

The values of normalized energy, greater than 1 for all the conditions, confirm the closed-loop nature of the error regulation from the participants, who exhibited higher dynamics than the target trajectory (Hanneton et al., 1997). The energy contained in the motion was significantly increased in the presence of sonification, however, no significant difference was found between the three feedback tested. This result confirms our hypothesis that continuous sonification can modify movement features. In the third experiment, the *Pointer* condition group tends to exhibit more energy than the *Target* group, either during training or retention test. Furthermore, the correlation between energy and performance grows stronger for the *Pointer* group in the retention test. This can indicate that learning to use extra energy to correct movement trajectory does not lead to dependency on the feedback, hence the larger improvement observed in the retention test for the *Pointer* group. The instantaneous and reactive aspects of the interaction is of course important to ensure the ‘ecological’ link between movement and perceived energy (Gaver, 1993b). The question of how this extra energy can be integrated without feedback dependency is still open.

Towards an auditory-driven proprioception?

The pointer sonification provides the participants with information on their own internal dynamics while performing the task. This information could be interpreted by the sensory system as an ‘augmentation’ of proprioception. The fact that participants interact with a tangible object in the tablet (the stylus) emphasizes the physical nature of the pointer feedback and the coherence with motion. The sensorimotor system could thus benefit from this richer sensory input. As a result, the sensory feedback prediction (forward model) could be faster and more accurate (Miall and Wolpert, 1996; Wolpert and Ghahramani, 2000). This mechanism is found in (Vogt et al., 2009) or (Schmitz and Effenberg, 2012): auditory feedback can be used to enhance the movement consciousness. The high energy measured in the participants’ movement in the pointer condition could be the manifestation of such a mechanism leading to a robust learning. In addition, positive reactions from the participants were found in the questionnaires. Recent studies examined the link between proprioception and motor learning. Rosenkranz and Rothwell (Rosenkranz and Rothwell, 2012) showed that integration and modulation of proprioceptive input induced a positive effect on motor learning. Wong et al. (Wong et al., 2012) showed that a proprioceptive training could reduce position error and increase speed, leading to a larger learning in an arm movement. These recent studies on sensory training support the theory of perceptual learning (Darainy et al., 2013) which describes sensory-motor learning entrained by perceptual changes. These results are in favor of our hypothesis, but further investigation is needed to address this particular question with continuous sonification.

Individual differences

None of the participants were excluded from the analysis based on performance criteria. The sometimes important variability we observed can be the consequence of a higher stimulation of the participants due to the sonification of their motion dynamics. This might have blurred some of the effects observed. We argue that some participants may be significantly less responsive to auditory-motor relationship, especially in this case where the attention is mainly focused by the visuo-manual regulation. We previously observed such a heterogeneity in the case of a closer auditory-motor relationship, under the paradigm of *sound-oriented task* (Boyer et al., 2014).

We believe the diversity in participants responses to sonic interaction is not addressed enough in the related literature. Modifying the sonification mapping through the experiment and according to the participant baseline performance could be considered (machine learning algorithms for multimodal mappings can offer this solution).

Conclusive remarks

The presence of continuous auditory feedback in a two-dimensional tracking task proved to have a positive effect on performance and learning. The three types of sonification led to a better tracking and had an impact on movement energy.

Future studies should address the 'physiological difficulty' of the task, meaning controlling the behavior of the target with different models (ballistic, two-third power law, bouncing on the sides). This would allow tuning the difficulty and evaluating performance at small time scales, and observing whether participants could learn a target behavior with the help of auditory feedback. Varying the spectral components of the target trajectory can also be a way to evaluate whether the phenomenons observed here operate at similar or different spectral ranges of motion. Adding or cutting higher frequencies in the target motion would make the trajectory respectively more or less difficult to follow, and allow to observe frequency-dependent correction mechanisms. The integration of the auditory feedback may also vary with the subsequent difficulty of the task.

To conclude, the results presented in this paper show the benefits of continuous sonification in a visuo-manual tracking task, highlighting the possibility to induce implicit learning by sonifying goal-related parameters or the participants' own motion. Our observations show that although error sonification can help performance, auditory feedback design for motor control should be also motivated by providing information on the performed action, relatively to the user. We showed the importance of using velocity-related modalities for continuous sonification, especially because of its potential link to energy in motion. We believe that continuous sonification of movement features should be further employed, especially considering interactive scenarios where increased energy is required. Physical rehabilitation, where the engagement of participants is crucial, can be an important application.

Acknowledgements This work has been funded by ANR French National Research Agency, under the ANR-Blanc program 2011 (Legos project ANR-11-BS02-012) and additional support from Cap Digital.

Chapter 6

Sonification for eye movement control

As illustrated in the previous chapter, tracking tasks are mainly studied with hand or arm movements. There is however a particular tracking movement that we produce continually, with very high dynamics, and that we are not much conscious of: the eye movement. In this chapter, we present an exploratory work on the sonification of eye movements dynamics to improve oculomotor control. Specifically, we focus on a specific action-perception coupling allowing to produce smooth eye pursuit in the absence of moving target.

In this chapter, we first introduce essential notions about eye movements and smooth pursuit, we then describe a visual illusion paradigm allowing to produce free smooth pursuit. A setup allowing kinematic eye parameter sonification and a dedicated training protocol are then tested in an experimental study. We report the main results regarding the sonification effects on performance and learning, as well as the main drawbacks of the chosen protocol, and emphasize important potential applications that should be addressed in this framework.

Note: this work is the product of a collaboration with the team of Jean Lorenceau and Arthur Portron at the Laboratoire des Systèmes Perceptifs (LSP) in Ecole Normale Supérieure, Paris. It is the first part of a series of experiments carried out jointly. We precise here that the field of this chapter is not central to this thesis, and that we address it carefully. Nevertheless, we chose to present here the motivations, protocol and preliminary results obtained, because of their originality and their potential for future work. We specifically think of developing eye writing interfaces for paralyzed patients. A second experiment is being developed at the time of writing this document.

6.1 Motivations

Eye movements are fundamental to both vision and motor control of the whole body. These movements allow to acquire visual information about the world surrounding us, but are also essential to anticipate, control, and plan our actions. Triggering intentional eye movements is thus mainly motivated by information acquisition. Moreover, eye movements are often driven by the perception of surrounding events, which can be sudden or unexpected. In that case, they allow for fast information acquisition necessary to attention and decision making (Krauzlis, 2005). This particularity of the action-perception loop thus makes eye movement control a distinctive paradigm for studying the coupling with auditory perception.

In this chapter, we are proposing to address auditory-motor coupling with eye pursuit movements. Several interests motivate this study context. Producing smooth pursuit eye movement relies on a moving target in the field of view (gaze following a car passing or a bird) (Lorenceau, 2012). In the absence of target, triggering so called ‘free’ eye pursuits is nevertheless possible, but requires a visual stimulus and a specific training, described in this chapter. As of today, there is no established protocol of learning free eye pursuit movements which meets general approval. The scientific community also lacks theoretical and experimental tools to evaluate and model inter-individual variability in this domain; this prevents from adapting learning protocols. The development of such techniques are among the key issues we investigate in this chapter. Some experts of oculomotor control, highly trained to smooth eye pursuit with reverse-phi, are indeed able to produce complex smooth trajectories, like cursive, writing with their eyes[1].

Developing real-time continuous sonification of eye movements for free smooth pursuit represents a technical and design challenge itself which remains to be met. These movements exhibit indeed very high dynamics, change rapidly and almost never cease. In addition, they might be often contaminated by involuntary movements (Krauzlis, 2005). Furthermore, the vision research literature do not mention continuous auditory-motor coupling for eye movement control.

Finally, the present work can benefit the development of writing interfaces or artistic devices based on eye movements. Specifically, we emphasize the benefits for paralyzed patients, either following a trauma or affected with neurodegenerative diseases like amyotrophic lateral sclerosis (ALS). Affected people suffer from severely reduced motor capacities, sometimes being able to control only their eyes.

In the following sections, we first remind the specificities of eye movements and especially eye pursuit. We also introduce a technique adapted and developed by Lorenceau (Lorenceau, 2012) allowing participants to initiate eye pursuits in the absence of moving target. We then propose a first experimental protocol in this context, which uses sonification for free pursuit movements learning. Results are presented on qualitative aspects and inter-individual

---

[1] `http://www.sciencesetavenir.fr/sante/20120730.OBS8484/quand-l-oeil-remplace-la-main-ecrire-avec-les-yeux.html`

differences. They allow to set the basis of a deeper investigation on the use of auditory feedback for eye movement control. This work presents a new experimental framework and a previously unseen utilization of sonification which, we believe so, deserves to be further explored.

6.2 Smooth pursuit eye movements

6.2.1 Eye movements

Let's distinguish here the different types of ocular movements on the basis of their function. First, smooth pursuit - one of the subjects of this study - allows the eye for following a moving target and to maintain incoming light rays from the target on the fovea, the most sensitive part of the retina, where visual acuity is maximum. This movement allows us to extract dynamical information that are central for visuo-motor actions (see chapter 5). Second, saccades (other topic of this chapter, which are a specific elementary movement) allow for fast exploration of a scene and rapid displacement to access a particular point of the field of view. Third, the eye extracts information during fixation periods as well, whether on the targeted spot or in the surroundings (peripheral vision).

These types of movement are generally voluntary. Reflex movements can also be produced by the oculomotor system. The optokinetic reflex triggers the pursuit of a target entering the field of view. The oculovestibular reflex allows to compensate relative head movements in order to keep the gaze focused on a precise point.

These functional movements are composed of elementary movements of the eyeball (from a kinematic point of view), which kinematic and dynamic characteristics distinguish them from other body movements:

- Drift is a relatively slow and smooth movement of the eyeball. A lower magnitude drift also exists, sometimes called micro-drift.
- Saccades are rapid movements of variable amplitude, allowing amongst others, to explore a visual scene quickly. They also appear when following a moving target ('catch-up saccades')[2], allowing the eye to correct rapidly its position towards a tracked target (analogous to arm saccades, see chapter 5).
- Micro-saccades, similar to the previous ones, exhibit lower amplitude.
- Tremor is a rapid oscillatory and low amplitude movement of the eyeball, which occurs particularly during fixations.

Saccades are the fastest supervised movement of the human body: the eyeball can reach spin velocities of several hundreds degrees per second. Besides, eye movements exhibit a

---

[2] Other types of saccades can be observed, varying with the context and goal, that we will not detail here.

large dynamic of speeds, which requires dedicated oculometric analysis and methods.

For the list to be as complete as possible, we should mention vergence, where both eyes move in opposite directions, converging or diverging. This movement allows for the adjustment of gaze position on objects located at various distances (along with accommodation which is a deformation of the eye lens).

6.2.2 Pursuit

Imagine a person standing still and following with his eyes an object moving to his right. The oculomotor system is activated and regulated so that the light coming from the object stays on the fovea. In order to ensure a smooth pursuit of the object, the pre-motor cortex gets information from the neurones in the visual dorsal areas about the ongoing trajectory. These neurons are sensitive to direction and velocity of the perceived motion. The background around the object then seems to move to the left: if gaze is maintained on the object, the background image moves in the opposite direction on the retina. In the absence of a target to follow (*i.e.* to initiate a 'free' pursuit movement), it is precisely this phenomenon of inverse perceived motion that prevents the establishment of a smooth pursuit and a continuous control. As a result, saccades appear, which aim at attempting to solve this ambiguity (Lisberger et al., 1987; Krauzlis, 2005; Lorenceau, 2012). Some studies demonstrate nonetheless that smooth pursuit eye movements can be produced in the absence of moving visual target. Ward and Morgan (Ward and Morgan, 1978) showed that it is possible to follow ones hand motion or sound targets in the dark. Berryhill et al. reported some examples of smooth pursuit following sound targets as well, or indicated by tactile or proprioceptive stimuli (Berryhill et al., 2006).

6.2.3 The "reverse-phi" illusion

There is another way to solve the perceived motion ambiguity, detailed by Lorenceau (Lorenceau, 2012) (see also cited references). The neurones sensitive to direction and velocity of the motion can inverse their preferential direction in the presence of a moving target if its contrast rapidly oscillates. This change causes the perception of an apparent motion in the direction opposite to the target one. This movement is called "reverse-phi motion". As detailed by Lorenceau, this motion is also perceivable when the eye initiates a voluntary *movement* in the presence of *fixed* targets, if their contrast oscillate around the background color (this oscillation of the stimulus is called 'flicker'). In this case, one can perceive a drifting sensation in the very direction of the eye motion. Figure 6.1, illustrates this reverse mechanism and typical stimuli used to trigger this apparent motion. A detailed example will be given in section 6.3.3. The apparent motion thus constitutes a positive sensory feedback of a smooth pursuit. This percept ("illusion") is then created by the very movement of the eye, establishing a perception-action loop. In the absence of voluntary movement from the observer, this percept is not visible, which precisely makes it, from a pedagogical point of view, delicate to explain to a subject.

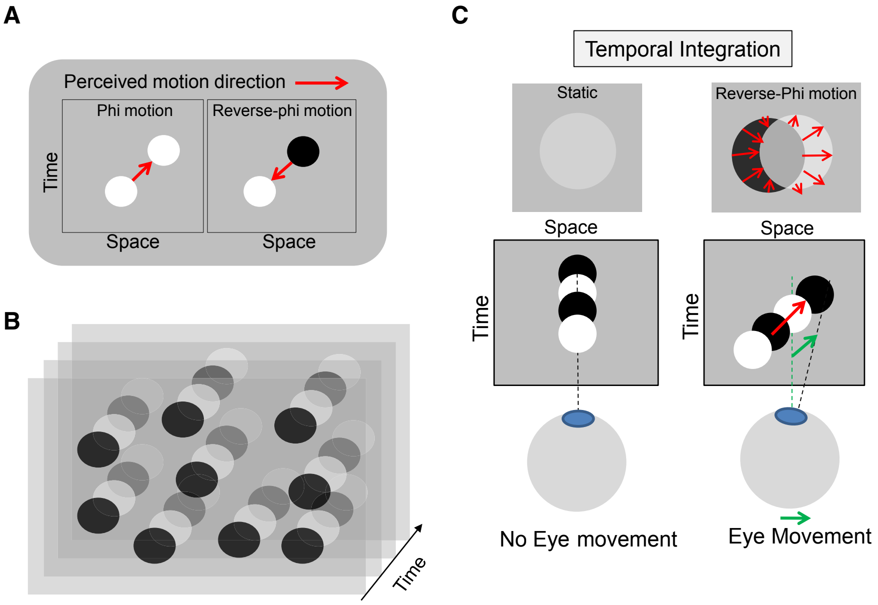


Figure 6.1 – Illustration of phi and reverse-phi motions in the spatiotemporal domain. A: left, a light point moves between two frames; creating the illusion of a motion towards the second position; right, the contrast polarity is reversed on the second frame, creating a perceived motion in the opposite direction - the reverse-phi illusion. B: typical example of stimuli used to sustain phi-motion induced smooth pursuit; pseudo-random pattern of static disks with flickering contrast around the background (∼3%). C: retinal spatiotemporal luminance when observing the previous stimulus: an eye movement produces the reverse-phi illusion on the retina leading to a congruent neuronal feedback, allowing to maintain smooth motion and to 'surf' on the stimulus image. Reproduced from (Lorenceau, 2012), courtesy of the author.

We propose to focus on the auditory-motor coupling in this particular framework, and to use similar sonification techniques to the those presented in this thesis, adapting our approach to the specific context of eye movements. The main objective of this work is to study the effects of eye movements kinematics sonification on perception and learning of voluntary reverse-phi motion (see details in section 6.2.5).

6.2.4 Auditory feedback in oculomotor studies

Thanks to the development of modern eyetrackers, and low-cost versions, various applications of eye movement control have been imagined. In the following paragraphs, we cite some of them which used auditory feedback. To our knowledge, no study described the usage of continuous sonification of eye movements for smooth pursuit learning.

Experimental studies did use sound as a stimulus, especially the ability of the auditory system to localize sound sources in surrounding space. Gauthier and Hofferer (Gauthier and Hofferer, 1976) and then Ward and Morgan (Ward and Morgan, 1978) observed that eye pursuit movements can be generated in the dark while following auditory targets moving in front of the subject. More recently, Berryhill et al. (Berryhill et al., 2006) compared different stimuli informing about the motion of a pendulum and measured the tracking gain (velocity) exhibited by the subjects. They compared auditory (loudspeaker attached to the pendulum), tactile (the experimenter moved the pendulum against the subject's arm) and proprioceptive (subjects moved the pendulum themselves) modalities. Results showed that tactile and proprioceptive stimuli provide more information for tracking that auditory stimulus, and led to higher tracking gain.

Baumann and Greenlee (Baumann and Greenlee, 2009) also used a moving auditory source but as a distractor[3] in 180° phase shift with the eye motion. They were able to measure an increased neuronal activity (posterior parietal cortex and lateral and medial frontal cortex) when the attention was divided between the two modalities. Besides, although they were submitted to a high cognitive load, subjects were still able to identify the movement of the auditory stimulus and their performance was not affected by the auditory distractor.

Kerzel et al. (Kerzel et al., 2010) have established a link between auditory perception and catch-up saccades. By observing the tracking gain diminution and the number of saccades produced after the brief and sudden appearance of disctractors while tracking a target, they showed that saccades can be suppressed during a short time after the appearance of a distractor. When the distractor was a loud and task-incongruent sound (10 ms white noise 'click' at 83dB(A)), the tracking gain was also less affected than with a visual distractor appearing at the periphery of the visual field.

It is interesting to note that sound has already been used to initiate smooth pursuit eye movements, but only as an external stimulus. But in the absence of any interactive aspect of the auditory feedback, motor control is driven one way, only from perception to action without closing the loop.

Finally, Madelain and Krauzlis (Madelain and Krauzlis, 2003) used pure tones (100 ms *beep*) to notify the presence of saccades on movements. The sound stopped clicking if the subject produced saccades; this represents a KP (knowledge of performance) feedback. Furthermore, a KR (knowledge of result) feedback is produced (2 *beeps*) in case of success in a trial, in addition to a video animation and money reward.

In the artistic domain, Hornof recently presented at the NIME conference (New Interfaces for Musical Expression) a short review (4 works) of musical control interfaces developed exploiting eye movement (Hornof, 2014). He pointed out that in spite of the great potential of these applications, many obstacles still restrain their development and use. Most of the interfaces presented use gaze as a pointer and a series of areas on a screen, triggering MIDI events or musical samples when gaze stays on them. Whether the controls are simple (sample triggering) or more complex (picking a tonality or interfacing with digital audio workstation and sequencer like Ableton Live), control is given to gaze position and not

---

[3] Neologism from P. Valéry: element capable of diverting thought on another object, *Tel quel II*, 1943, p.237

to eye dynamics. They indeed propose discrete and not continuous actions and lack of musical expressiveness. Finally, specific features of the oculuometric signal (blinks, noise, tremor, high velocities and accelerations, etc.) make designing and learning very difficult with these interfaces.

6.2.5 Hypothesis

Similarly to the visuo-manual tracking task (see chapter 5), we hypothesize that continuous auditory feedback imaging movement dynamics can benefit learning of smooth eye pursuit production.

Specifically, we suggest that eye movement sonification can enhance proprioceptive feedback during motion. Indeed, although it has been shown that proprioceptive feedback is available in eye motion, it is integrated unconsciously (Steinbach, 1987). Some authors even argue that this sensory feedback is not coded in eyes position coordinates in their orbit, but relatively to their motion in space, integrating head movements (Wertheim, 2010). In addition, as every sensory feedback, this information comes to the central nervous system with delay and noise. Bridgeman (Bridgeman, 1995) insists on the importance of efference copies (neural images of motor commands used in the regulation of sensorimotor loops) for proprioception as an extra-retinal information source. This lack of proprioceptive feedback is supposed to be detrimental to free smooth eye movements with reverse-phi, as the necessary visual percept itself is conditioned to initiated movements. Besides, the few experiments carried out on learning free reverse-phi pursuit (actually in preparation with the associated team on this work) are quite ineffective, particularly due to the lack of (proprioceptive) feedback.

The dynamical nature of the reverse-phi illusion implies for the subjects to 'see' the illusion in order to understand it and control it (the eye must move in order to get perceived motion, which is a quite unusual sensation). For a beginner (who does not understand the illusion yet), the lack of proprioceptive feedback could prevent to know when smooth pursuits are initiated and lead to poor inverse models for the sensorimotor system. The sonification of smooth motion versus saccades could act as a KP sensory feedback, helping the subjects to identify saccades but also smooth pursuits that are being initiated.

6.3 Sonification for free smooth pursuit eye movement learning

6.3.1 Introduction

This experimental work proposes the use of an auditory feedback to help learning 2D free pursuit eye movements with reverse-phi stimuli. Particularly, we propose a system that enables the sonification of saccades and initiated free pursuit movements in real-time. A training protocol has been developed using 7 geometric patterns and will be evaluated in two conditions: a tracking task of a target drawing the patterns, and a free condition where subjects have to reproduce the patterns with free smooth pursuit using a reverse-phi stimulus.

6.3.2 Participants

Fourteen participants volunteered for this experiment, and twelve finally participated and followed the complete protocol. The nominal experiment last for 9 daily sessions of one hour each. It took place in the LPC lab of Ecole Normale Supérieure of Paris. Two subjects came from this lab but were nonetheless naive regarding our protocol. All participants signed an informed consent and were paid for their time. Six out of the twelve subjects had already participated in experiment involving eyetracking. One subject (#10, included on purpose in the study) had already performed a reverse-phi motion training one year prior to our experiment for about ten hours; this case will be detailed later.

6.3.3 Setup and stimuli

Equipment

The experiment took place in a soundproof booth. Eyetracking was achieved with an Eyelink 1000[4] eyetracker. It is composed of a camera placed on a table facing the subjects and a dedicated computer (first machine). An infrared light is pointed towards the subjects' eye (the tracking is monocular) and the reflection on the retina and the cornea is captured by the camera. Gaze position is computed integrating the apparent shape of the retina (a circle when the eye looks straight, changing to an ellipse away from the center) and the cornea reflection. A five point calibration is performed to provide absolute position of gaze on the computer screen. The technical data sheet of the eyetracker is available in appendix C.

Visual stimuli were presented on a 1024x768 pixels operating screen (51,3 x 32,1 cm) facing the subjects. The eyetracker is placed underneath the screen, at 57 cm from the subjects' eyes (distance is ensure by placing subjects' chin on a chin piece). From that distance, a 1 cm gaze displacement on the screen corresponds to a rotation of 1° of the eye in his orbit. Data recording includes (x, y) gaze position in the screen space and pupil diameter, at a 500 Hz sampling frequency. The calibration is consistently run before each recording (it has to be repeated each time subjects move their head out of the chin piece). This allows the eyetracker to compute gaze position in the coordinate system of the screen in pixel units (the same as the stimuli). In this 'head free' configuration (subjects don't wear any device on their head) subjects are instructed to move their head as little as possible during the recordings.

---

[4] `www.sr-research.com`

A second computer (HP, Intel Core i7, Windows 7) commands the eyetracker for recording and generates the visual stimuli (targets and reverse-phi pattern). A third computer (MacBook Pro, Intel Core 2 Duo, OSX 10.8) receives oculometric data as well at 250 Hz and creates the sonification in real-time.

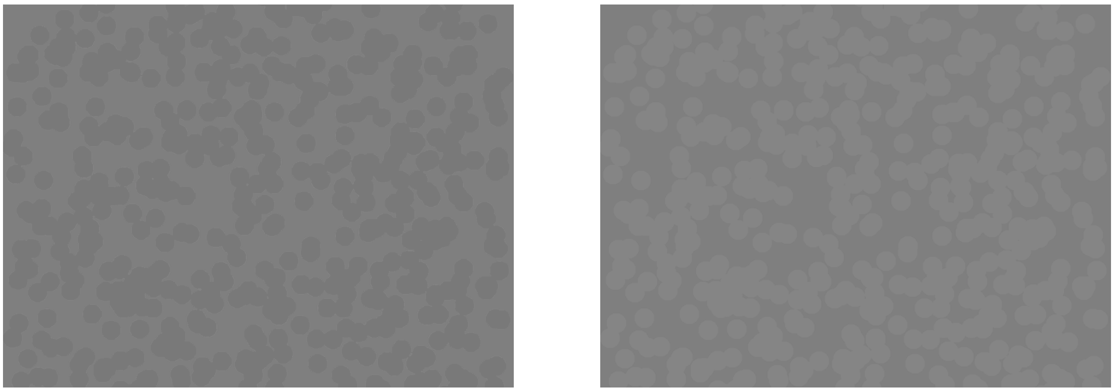
Visual stimuli

The visual stimuli, target, and reverse-phi patterns are generated by the second computer and displayed on the screen at 60 Hz. The reverse-phi stimulus is dynamic: the light/dark disks are renewed in an other random pattern every 50 frames, in order to avoid possible gaze fixation on the shapes they may form. The stimulus is generated on the entire screen surface with the following parameters:

- 500 disks, pseudo-randomly scattered (in order to avoid large clusters)
- disks diameter: 40 pixels (2°)
- disks are renewed after 50 frames (833 ms)
- gray palette:
  - ○ dark disks: RGB(121,121,121), luminance 11.490 cd.m^{-2}
  - ○ light disks: RGB(133,133,133), luminance 14.665 cd.m^{-2}
  - ○ background: RGB(127,127,127), luminance 13.018 cd.m^{-2}
- flicker: 10 Hz.

Figure 6.2 shows an example of two frames which compose the reverse-phi stimuli, and alternate every 100 ms (light and dark contrast polarity, the background stays the same).

The moving target drawing the patterns is represented by a light gray disk of 10 pixels diameter (0.5°) with RGB color (165,165,165), being 25.580 cd.m^{-2} luminance.



(a) Dark contrast

(b) Light contrast

Figure 6.2 – Reverse-phi stimulus for the experiment. Contrasted images alternate at a rate of 10 Hz, hence a contrast oscillation of the disks between 11.490 cd.m^{-2} (a) and 14.665 cd.m^{-2} (b). Background color is identical on both images.

Sonification

A custom program has been developed to transmit the eyetracker data in real-time with OpenSoundControl protocol[5] to the third computer. The program uses the eyetracker built-in API and polls data from its own machine. A UDP connection protocol allows routing the data to the third computer. Incoming data are then processed with a custom program built under the Max/MSP environment[6].

Two distinct sonifications are proposed: one for smooth pursuit eye movements and one for saccades (see figure 6.3). The program allows for the sonification and interprets the messages from computer 2 (record start/stop, eye presence confirmation, shut down of audio synthesis after one second of inactivity). A screenshot of the sonification interface is shown figure 6.4; the commands and data received from computer n°2 are visible ('Input') as well as filtering parameters ('Controls') and raw signal waveforms of continuous speed and saccades. Two different synthesis engines generate pursuit and saccade sounds, from horizontal and vertical gaze speed signals. Specifically, pursuit sounds are generated from the squared norm of tangential velocity gaze vector. This signal is then filtered by a 20 samples median filter (Bevilacqua et al., 2005). It commands a resonant filter (Max object *reson~* factor Q=10), driven between 100 and 500 Hz, operating on a pink noise. The low end of the spectrum is then cut off for clarity (figure 6.3). From an ecological point of view, the sound of smooth pursuit has been designed to evoke the sound of wind, or rubbing a surface. It insists on smoothness and continuity metaphors as well as respecting the dynamic range of the motion.

Saccade sounds are generated from the gaze acceleration signals (figure 6.3). After computing the squared norm of the acceleration vector, the signal is slightly smooth logarithmically. The resulting signal commands a monopole low-pass filter between 400 and 1000 Hz filtering pink noise (the low end of spectrum is also cut for more clarity). The envelope of the sound is then shaped with a 5 ms up and a 500 ms linear down-ramp. If the velocity of the eye exceeds 100°/s the saccade sound is triggered, and illustrates the profile of the saccade that is being produced. Below this threshold, no saccade sound is produced. This threshold velocity is recognized as the upper velocity range of possible pursuit production for humans (Meyer et al., 1985). The method of saccade sonification is not based on the simple triggering of an audio event. The intensity and the acceleration temporal profile of the saccades are preserved and included in the auditory feedback. In this way, the system we developed enables the subjects to perceive the intensity of the saccades they produce along a continuum, both through loudness and spectral content of the sound. Meanwhile, the pursuit sound is turned off when a saccade occurs (using a 50 ms up and 100 ms down linear ramp) in order not to receive both feedback simultaneously; as it impossible to produce both pursuit movement and saccades at the same time.

---

[5] `www.http://opensoundcontrol.org/`

[6] `www.cycling74.com`

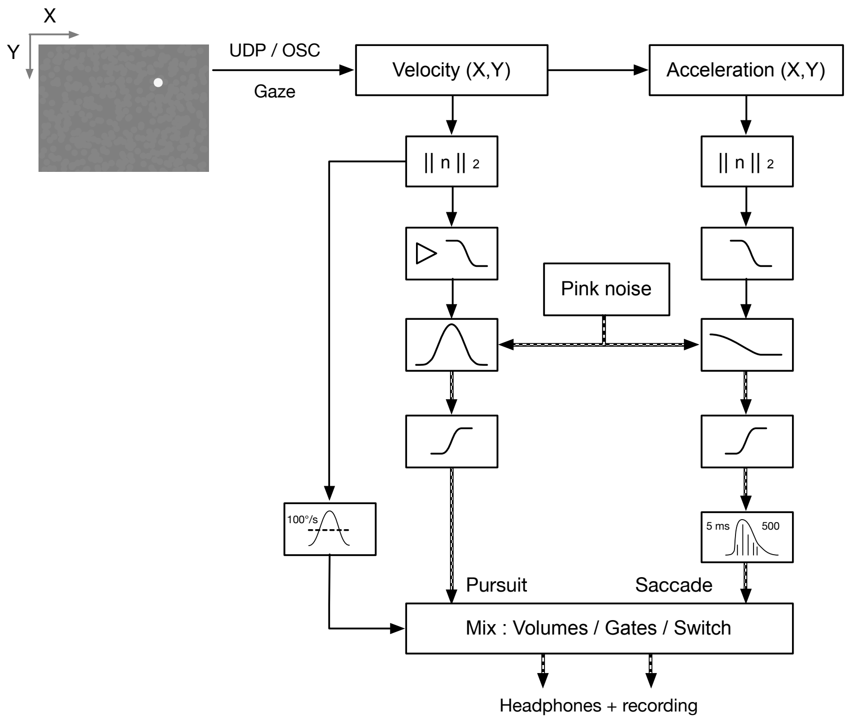
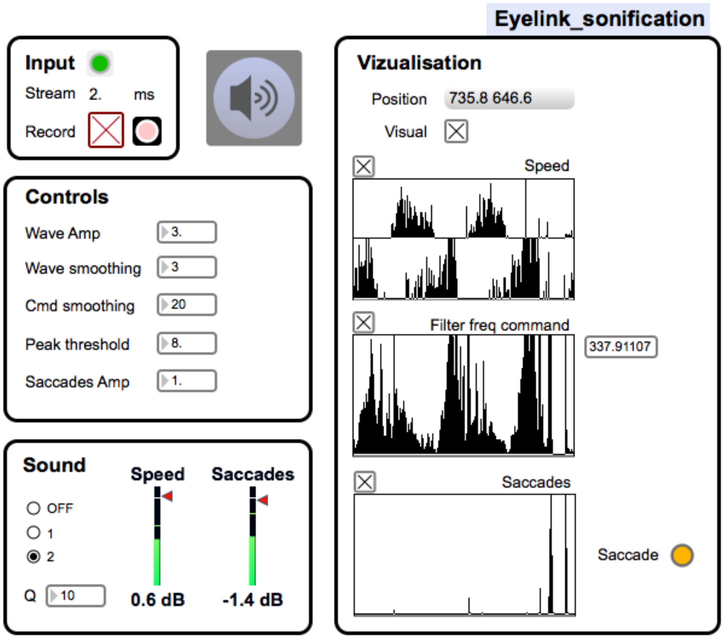


Figure 6.3 – Schematic of the pursuit and saccade sonification processes.



Stream 2. ms

Wave Amp 3.

Wave smoothing 3

Cmd smoothing 20

Peak threshold 8.

Saccades Amp 1.

○ OFF
○ 1
● 2

Q 10

Speed 0.6 dB

Saccades -1.4 dB

Position 735.8 646.6

Visual ☒

337.91107

Figure 6.4 – Screenshot of the Max patch presentation interface, receiving the eyetracker data and generating the sound feedback on pursuits and saccades.

6.3.4 Protocol

The experimental protocol includes a preliminary session, allowing us to detect subjects who are uncomfortable with eyetracking measurements - which can be relatively tiring -

or who present a very bad tracking ability. This session also allowed us to check if the subjects perceived correctly both sonification aspects. Two subjects have been withdrawn from the panel following the preliminary session.

A first tracking test is performed on the 7 patterns of the corpus used in the training sessions. Subjects are asked to follow a mover drawing each pattern twice in a row. Each pattern is randomly repeated 5 times, for a total of 35 trials during the preliminary session. In a second time, the experimenter describes to the subjects the notions of smooth eye pursuits and saccades. Two audio examples of sonified movements are then presented to the subjects, one containing much smooth pursuit and small catch-up saccades, and one containing many saccades. They are asked to determine which recording was the most successful under the criteria of the experiment which is to produce trajectories as smooth as possible. Finally, subjects are introduced with the reverse-phi stimulus. The experimenter invites them to slowly move their head freely, and try to perceive the 'sliding' resulting motion. This motion should be correlated in amplitude and direction with the head movement. They are also encouraged to accommodate their vision beyond the screen, which can usually help to perceive the reverse-phi illusion. They are given the objective to find this percept again during the training sessions, but with eye movement only.

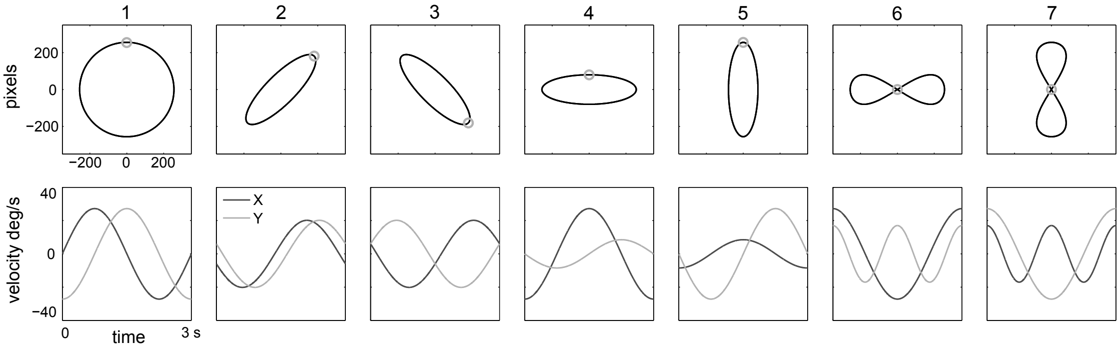


Figure 6.5 – Corpus of tracking patterns and associated velocity profiles (bottom). The starting point of the pattern is indicated. Two 'laps' of each pattern are performed in the 'Follow' phase.

The training protocol is scheduled over 8 daily sessions; four are performed with the sonification, and four without. Subjects were randomly assigned to two groups: group 1 started the sessions with sonification and finished with four sessions without. Group 2 did the opposite. Each session is composed of 4 blocks of 5 repetitions of each pattern, randomly presented in the block (for a total of 140 trials per sessions). Performing one block of test takes approximately 12 minutes. A five minutes break is taken between each block. The complete experimental plan is therefore 8x4x7x5 = 1120 trials.

The target patterns represent a circle, four ellipses and two figure eight. Figure 6.5 shows the trajectories and their velocity profiles. Patterns are inscribed in a 512x512 pixels square and centered on the screen. The ellipses have an eccentricity of 0.95 and are either vertical, horizontal or shifted 45°. The figures eight are scaled Lissajous curves, of parameter $N = 2$, either vertical or horizontal.

A trial is composed of two phases, which have to be distinguished: the 'Follow' phase during which subjects have to track the moving target along the patterns, and the 'Trial' phase, where subjects have to reproduce the patterns with free smooth eye pursuit (see figure 6.6). Each trial starts with a priming of the pattern with the starting point indicated. Subjects then fix the starting point for 1 second before it draws the pattern in 3 seconds per lap (two laps). In the next phase, the priming is presented and subjects have then 6 seconds to reproduce it with smooth pursuit using the reverse-phi stimulus. Each trial is then composed of a training and a performance phase. The reverse-phi stimulus is continuously present in both phases, so that subjects are exposed to constant conditions in both phases. The quality of the reproduction in the Trial phase is mainly the quantity of smooth pursuits produced; subjects are told that they can draw smaller or shifted pattern if they prefer, but that they can practice with the model during the Follow phase.

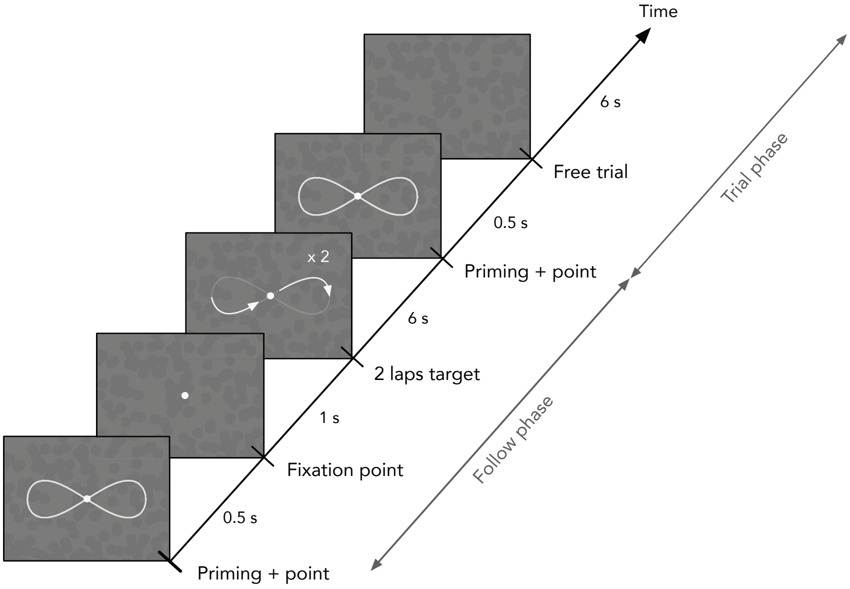


Figure 6.6 – Scheduling of each trial, divided into the 'Follow' phase (following the moving target) and the 'Trial' phase (free pursuit production) with the reverse-phi only.

Two questionnaires are proposed to the subjects at the end of each session. A first one asks them about their visual perception consciousness, their habits with visual environment and about light symptoms they may present (visual fatigue, regular headaches, eye stinging, etc.). This questionnaire is filled once in the preliminary session, and they are asked to mark evolution of these criteria after each training session. The second questionnaire allows the subjects to rate their performance after each bloc of training and for each pattern, answering the question "Do you think you reproduced the pattern smoothly?" on a 0-10 scale (meaning 'Not at all' to 'Perfectly').

6.3.5 Data analysis

Segmentation

The oculometric data are low-pass filtered at 100 Hz with a Gaussian filter to eliminate noise. A saccade is identified if the instantaneous eye velocity exceeds 100°/s (same value as the sonification threshold). The time segments around saccade movements are counted and analyzed: peak and average velocity, distance traveled and time. Segments with no saccade are examined to look for smooth pursuits. Accumulated eye displacements no larger than 2.5° between two saccades are excluded. Fixations are detected if the horizontal and vertical spatial distributions of the displacement on a segment exhibit a standard deviation smaller than 30% of the 2.5° distance threshold. Smooth pursuits should consequently match the three kinematic criteria of speed, produced displacement and standard deviation of spatial distribution. Saccades and pursuit segments no larger than 3 points (6 ms) are also excluded as they are taken for measurement artifacts. Blinks are recognized as short fixations points with $(0, 0)$ position value (meaning no eye detection).

Performance

During the Follow phase, performance is mainly assessed by the pursuit rate (the cumulative duration of smooth pursuits over the total trial time, 6 seconds), and the precision of the tracking. As the tracking can be quite inaccurate if the subject moved his head during the recording, refracts may occur while computing spatial precision. Some subjects exhibited this behavior, especially by the end of the blocks when fatigue was more important. We thus focus on velocity errors to evaluate the tracking precision. Specifically, the velocity gain is measured on the pursuit segments validated by the detection algorithm. The number of saccades counted during a trial is evaluated, knowing that the general and expected behavior is an alternation of saccades and smooth pursuits or fixations.

During the Trial phase, the main evaluation criteria are the number of initiated pursuits and the time ratio they represent. However, it appeared that few subjects were actually able to produce and control their pursuit movements enough, making not possible to assess the quality of spatial reproduction of the pattern. The evaluation of the effect of sonification in this phase proved difficult.

6.3.6 Results

Subjects exhibited a large inter-individual variability, for all the evaluated parameters. The duration of the trials (6 seconds for Follow, 6 seconds for Trial) participates as well to the variability. In this section, we first present observations on the tracking performance, especially in the Follow phase, however, without statistically significant conclusions when averaging the subjects (only tendencies are observable among groups). Then, specific behaviors are analyzed by taking the examples of four particular subjects, #6, 8, 10 and 12.

Follow phase: target tracking

Figure 6.7 shows the number of saccades produced per trials, as a function of pursuit rates for all subjects in Follow phase. A clear inverse linear relationship is visible (this occurs with all the patterns this time). This results was somehow expected; in order to increase the pursuit rate, subjects necessarily have to reduce the number of saccades during this phase. We will see that this correlation does not stand any more in free conditions.

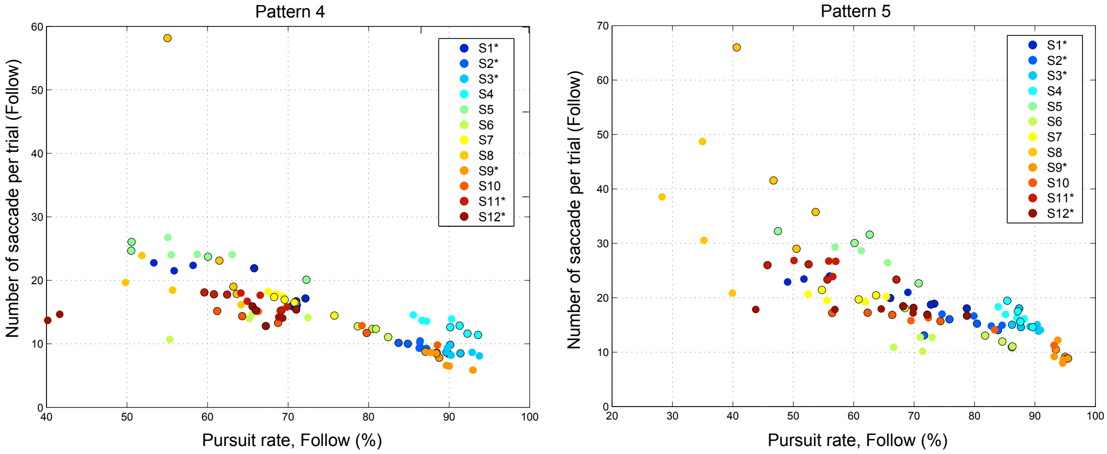
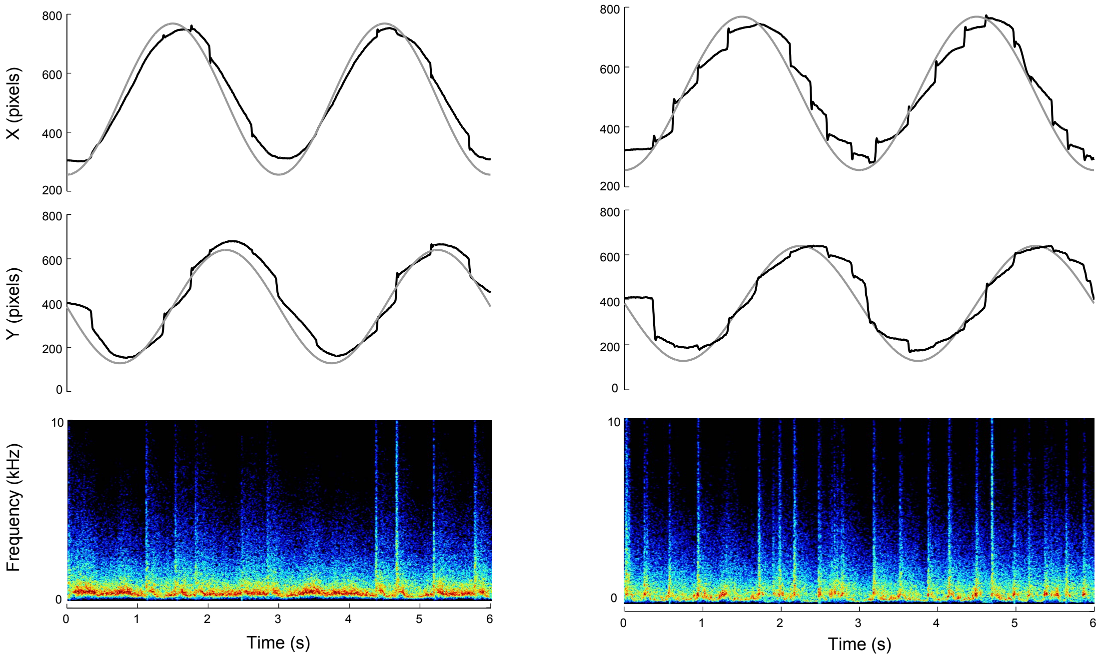


Figure 6.7 – Saccades count versus pursuit rate for patterns 4 (left) and 5 (right) in Follow phase. Each point is the value averaged over a session (8 points per subject). Black rimed points indicate sessions performed with sonification. Subjects from group 2 are indicated with '*' in the legend. Distributions show an inverse linear correlation between the two parameters.

The typical behavior of the subjects exhibited, as expected, a succession of smooth pursuits, of various lengths, interrupted by catch-up saccades. An example of recorded trajectories for pattern 1 (circle) is presented on figure 6.8. The spectrogram of the subsequent sonification is also presented. Saccades are clearly visible on the trajectory (black line, target is in gray) and on the spectrogram as well, represented by short and wide range spikes. The left example illustrates a nominal subject, with small catch-up saccades and a high level of pursuit sound, visible in the lower end of the spectrogram. On the right, the subject exhibits larger and more energetic saccades, which are clearly visible on the spectrogram. As a result, far less pursuit sound is produced (not visible in the low frequencies). Both examples illustrate a tracking delay between gaze and the target trajectory.

A large variability between the subjects can already be observed on qualitative assessments. Some subjects produced hardly recognizable patterns, mainly due to numerous blinks or important loss of concentration.

An analysis of the pursuit rate has been performed between the two groups of subjects.



(a) Nominal subject (b) Less successful subject

Figure 6.8 – Examples of gaze trajectories in Follow phase for pattern 1 (circle, gray line) and spectrograms of the subsequent sound produced. Saccades cause severe irregularities in slopes of the gaze and wide range spikes in the spectrogram. Pursuit sound is visible in the low frequency range on the left (subject #4). On the right (subject #5), tracking is degraded with more saccades of higher amplitude.

Figure 6.9 shows the pursuit rate for pattern 3 for both groups. Values are averaged across pattern repetition (5) and blocks (4) in each session. Although there were no significant effect neither of groups nor of the sessions, it seems that the arrival of sonification in the 5th session helped group 1 subjects to produce longer pursuits. The reduction of the variability by the end of the experiment tends to confirm this observation. Besides, group 1 shows no learning between sessions 1 and 4.

As far as group 2 is concerned, a small increase of pursuit rate is visible during the first 4 sessions. The pursuit rate starts decreasing after sonification has been removed, and goes down to 72%, which is below the level of the session 1. In spite of a little consolidation of the performance in the first sessions without sound (5 and 6), the improvement initially observed do not seem robust to sonification removal. Variability increases by the end of the experiment, indicating that some subjects could however maintain higher performances.

It can also be noticed that the performance levels for the two groups are slightly different. This illustrates a certain heterogeneity between them and points the issue of normalizing performance (individual evaluation may be necessary). Surprisingly, some subjects exhibited very low spontaneous pursuit rate in the Follow phase (down to 50%). This indicates that the exclusion criterion on tracking capacities in the preliminary session should be more selective than it was.

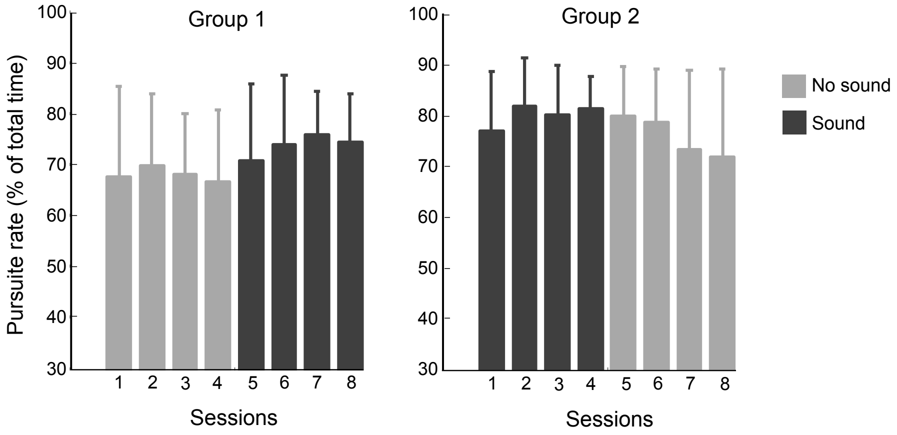
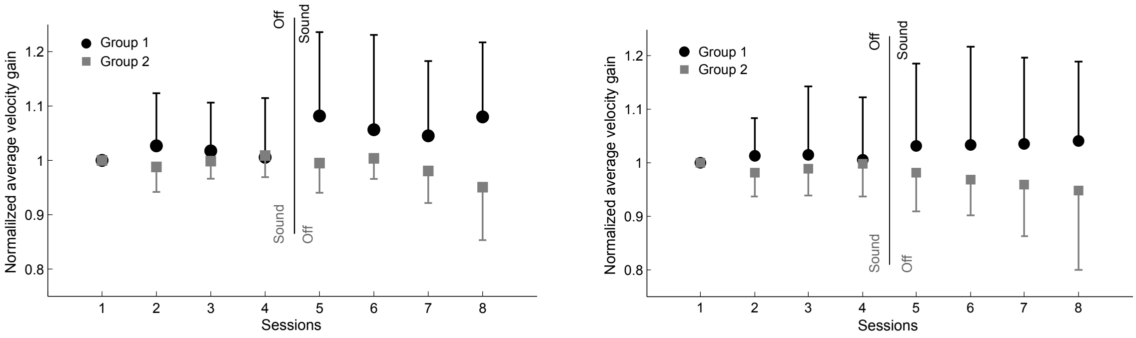


Figure 6.9 – Pursuit rates for pattern 3 in Follow phase for both groups. Error bars indicate 95% confidence intervals. Statistical analysis shows no significant effect between the sessions.

The evolution of the pursuit gain is also affect by the variability between the subjects. In spite of a large distribution, the effects of the feedback condition change are observable. Figure 6.10 shows the evolution of gain since the first session for patterns 3 and 5. The condition change in session 5 led to a small increase of gain for group 1 (now receiving the sonification) and a decrease for group 2. This tends to confirm that learning for group 1 has been helped by sound, and that group 2 exhibited a non robust learning, even maybe a feedback dependency. As no statistical significance is found, individual analysis would be required (see next sections). Furthermore, results observed in pursuit rate or gain largely depend on the targeted pattern; the observations described previously are not visible for all the patterns.



(a) Pattern 3

(b) Pattern 5

Figure 6.10 – Evolution from session 1 of the pursuit gain for patterns 3 and 5 in Follow phase. Group 1 started without sonification (dark points), group 2 (light squares) did the opposite. Error bars indicate 95% confidence intervals.

Trial phase: free pursuit generation

In spite of the quick reverse-phi stimulus familiarization and quasi inexistent experience of the subjects to control free pursuit with it, nine subjects out of twelve managed to produce smooth eye pursuit movements without target. This result is within the average observed in the few prototypical studies carried out by the LSP team. The three remaining subjects produced, on average, less than 10 validated pursuits per session and per pattern (that is over 20 trials of 6 seconds per pattern). In contrast, the spatial reproduction of the patterns is not achieved in the Trial phase.

A high level of performance in the Follow phase did not necessary lead to the same success in the Trial phase. Conversely, some subjects who exhibited poor tracking abilities in the Follow phase did produce an important number of free smooth pursuits. Figure 6.11 shows the distribution of pursuit rates for Trial versus Follow phase, for patterns 1 and 7 (using the same color codes as previously). Pursuit rates in Trial phase include only the produced and validated pursuits; if none were produced, the rate is 0. It appears clearly that success in Follow phase is not correlated with free pursuits production. Subject #4, for instance, was one of the best follower but produced very few free pursuits. This illustrates that predicting the ability of a subject to produced free pursuits with the reverse-phi stimulus should take other factors into account than the tracking performance. This figure also illustrates the upper limit of pursuit rates with the moving target: 95% for the circle, 90% for the horizontal ‘8’, highest scores, performed by subject #9. In Trial phase, the rates hardly exceed 50% (maximum for subject #7).

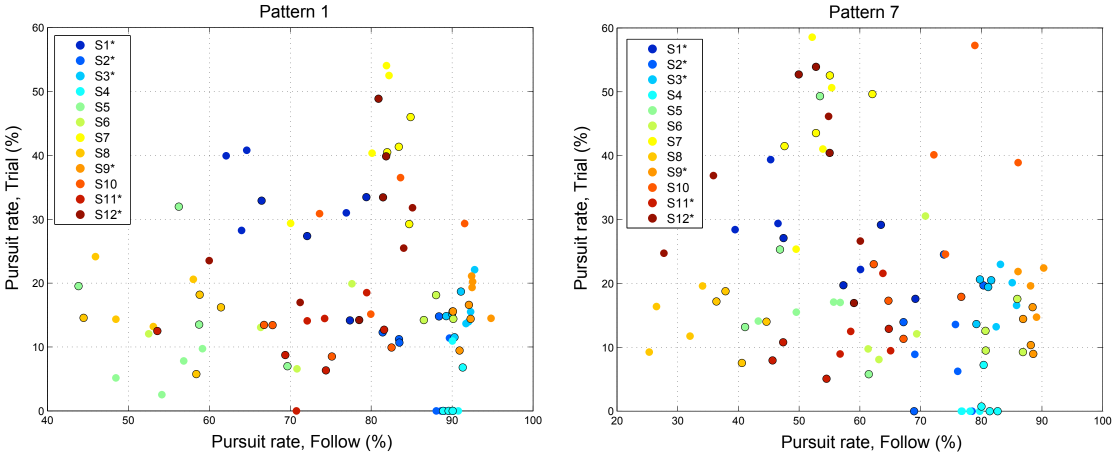


Figure 6.11 – Pursuit rates in Trial phase versus in Follow phase, for patterns 1 and 7. Each point is the value averaged over a session (8 points per subject). Black rimed points indicate sessions performed with sonification. Subjects from group 2 are indicated with ‘*’ in the legend. Distributions indicate no clear relationship between pursuit production in Follow and Trial phases.

It is worth noticing that patterns are not entirely reproduced during the first sessions. Subjects generally exhibit a spatial drawing much slower than the target did in the Follow phase: they seem to focus on initiating smooth pursuits, no matter the timing and the overall trajectory produced. The perception of the reverse-phi illusion and of the sonification may have focused much of the subjects' attention. By the end of the sessions, subjects achieved to produce two laps of the patterns, but often at the cost of pursuit quality.

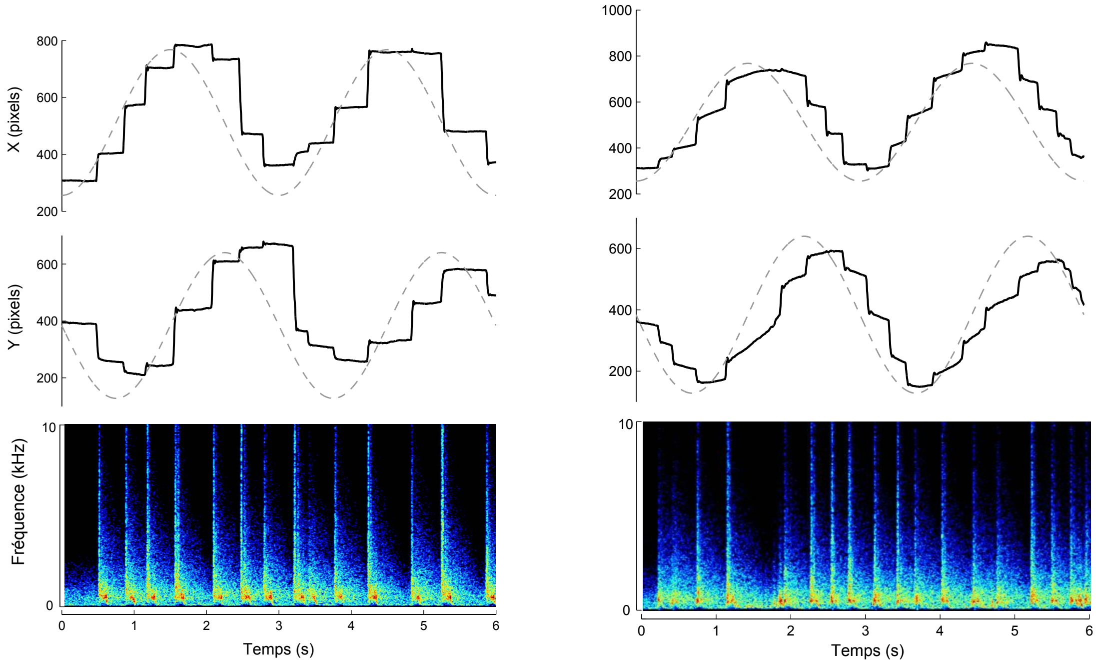
It was nonetheless expected that subjects could not perfectly control the speed of their free pursuits (whereas averaged velocities in the Follow phase are regulated by the velocity of the target). This is indeed one of the main challenges of learning free pursuit movement control, even with extensive training. The average velocity of smooth pursuits detected in this phase was $12.5 \pm 9.2°$/sec. The fact that both closed loop and open loop pursuits are mixed in the trials might explain the large standard deviation. As a result, this average velocity certainly differs from the 'theoretical' control velocity of the free pursuits (which is thought to depend on the reverse-phi parameters). This value would be measurable if subjects could maintain and control the pursuit motion with the reverse-phi stimulus.

Two examples of trajectories recorded in the Trial phase are presented on figure 6.12. The left side shows the example of a subject who did not produce pursuit movements; the resulting trajectory is a series fixation points (approximately half a second long), alternating with large amplitude saccades, with which the subjects try to draw the global shape of the original pattern (dashed gray line). As the subject is unable to move his eyes smoothly, he attempts to represent the global shape by moving with large saccades. This behavior is typical of subjects who were not able to learn the task, although they seem to understand clearly what is asked to them. The second example, on the right side of figure 6.12, shows trajectories with non-zero slopes, indicating significant displacement apart from saccades, which are smooth pursuits. The displacements during the saccades are also less important in this example. It can be noticed that, in this phase, the presence of smooth pursuits is not necessarily correlated to a decrease in the number of saccades.

Effect of sonification: individual behaviors

As we saw previously, very heterogeneous behaviors are observed among the subjects, and in both phases. In order to assess the effect of movement sonification on each subject and identify typical strategies or behaviors, we propose to compare individually all the trials performed with and without sonification. The same performance parameters than previously are used. For each pattern, subjects had 80 trials with and 80 trials without sonification (5 repetitions over 4 blocks, during 4 sessions). We compare these series of 80 trials with Student pair tests (*t-test*, 95%) for each subject and pattern. This method allows to avoid a normalization of each subject performance. Learning effects within the 4 sessions per condition are not accessible in this case. The effects of sonification are first evaluated without specifying the alternative hypothesis (no prior about the evolution of averages). In a second time, we specify the tail according to the main hypothesis of this experiment:

- the presence of auditory feedback favors the pursuit rates during the Follow phase;



(a) No pursuit

(b) Free pursuits initiated

Figure 6.12 – Examples of gaze trajectories in the Trial phase for pattern 1 (circle), and spectrograms of the subsequent sound produced. The dashed line represent the trajectory of the pattern (no target in this phase). Left (subject #4), no pursuit are produced: the displacements are caused by successive saccades with fixation points. Right (subject #1), subject initiates smooth pursuits and produces smaller saccades.

- the presence of auditory feedback can increase velocity gain during the Follow phase;
- the presence of auditory feedback favors the apparition of free smooth pursuits during the Trial phase.

The three tables 6.1, 6.2 and 6.3, contain the test significance results for the three hypothesizes. The blue color corresponds to a significant effect consistent with the hypothesis, whereas the red p-values indicate opposite significance. Individual behaviors can therefore be identified regarding our hypothesizes.

Table 6.1 – Effects of sonification: *p-values* for pursuit rate in Follow phase.
$** = p < 0.05$; $* = p < 0.1$, '$\cdot$'= non significant.

	Subjects											
Pattern	1	2	3	4	5	6	7	8	9	10	11	12
1	**	**	**	.	.	**	**	*	**	**	*	**
2	**	**	**	**	.	**	*	**	.	**	**	**
3	**	.	.	**	.	**	.	**	.	**	.	**
4	**	.	.	**	.	**	.	**	**	**	.	**
5	**	.	**	.	.	**	.	**	.	**	.	**
6	**	**	**	**	.	**	.	**	.	**	*	**
7	**	.	**	**	.	**	.	**	.	**	**	**

The results illustrate as well the statistical heterogeneity observed, between subjects and within target patterns. Investigating movement features as a function of target patterns

Table 6.2 – Effects of sonification: *p-values*for velocity gain in Follow phase. $** = p < 0.05$; $* = p < 0.1$, ‘·’= non significant.

	Subjects											
Pattern	1	2	3	4	5	6	7	8	9	10	11	12
1	**	**	**	**	.	**	.	**	.	**	.	**
2	**	**	**	**	.	**	.	**	.	**	.	.
3	**	.	**	**	**	**	.	**	*	**	**	.
4	**	.	.	.	.	**	.	**	.	**	**	**
5	**	**	**	.	.	**	.	.	**	**	**	**
6	**	**	**	**	.	**	.	**	**	**	.	.
7	**	.	**	**	.	**	**	**	.	**	.	.

Table 6.3 – Effects of sonification: *p-values* free pursuits count in Trial phase. $** = p < 0.05$; $* = p < 0.1$, ‘·’= non significant.

	Subjects											
Pattern	1	2	3	4	5	6	7	8	9	10	11	12
1	**	**	.	.	**	.	.	**	**	**	**	**
2	**	.	.	**	**	**	**	**	**	**	**	**
3	**	**	.	.	**	**	.	**	**	**	**	**
4	.	**	.	.	**	.	**	**	**	**	.	**
5	.	**	.	.	**	**	**	.	**	**	**	**
6		*	.	.	**	*	**	**	**	**	**	**
7	**	**	.	.	**	**	.	**	**	**	.	**

would require to know the specific dynamics of free pursuit movements they require, also depending on the reverse-phi stimulus. Patterns have been empirically ranked as "easy" (circle), "medium" (ellipses), and "difficult" (figure eight). The circle is executed with a constant tangential velocity and curvature, ellipses include larger curvatures, and both figures eight include an inflection point. As far as we know, there is no established model of oculomotor control of free reverse-phi pursuit movement. As a result, no difficulty function was available to normalize pattern difficulty in the experiment. We considered then the various patterns as different trials with their own specificities. Results indicate that patterns add another parameter to performance variability, which should probably depend also on individuals. Interestingly, the relative performance levels between patterns are nevertheless quite consistent for every subject along the experiment (see below).

In the following sections, we propose to detail the motor performances of four subjects, who exhibited specific behaviors, consistent or not with our hypothesizes. These subjects showed significant differences in performance between the sonification and no sound conditions in tables 6.1, 6.2 and 6.3. Learning behaviors are also identified. Firstly, subject #6 illustrates a positive reaction to auditory feedback in Follow phase, but a negative one during free trials. Subjects #8 and 12 exhibited the most important - and positive - effects of sonification in both groups. They enable the observation of learning and retention effects due to sonification presence at the beginning or at the end of the training protocol. Finally, the case of subject #10 is described, who already received a training on free pursuits production with a reverse-phi stimulus, during a slightly different protocol. His - quite unexpected results - revealed important observations on adaptation and re-learning mechanisms in this experimental context.

Case of subject #6 This subject (female, aged 28) was assigned to group 1, she thus started the experiment without sonification. During the first four sessions, she exhibited a gradual improvement of tracking motion in Follow phase (visible on figure 6.13), in terms of both duration and quality. From session 5, the subject received sonification. From this session, performance suddenly increased to reach a plateau phase, stable until the end of the experiment. Statistical analysis revealed a significant effect of sound for all the patterns of the corpus on pursuit rate ($p < 0.0001$) and gain ($p < 0.0001$). Looking at the learning curve in the 4 sessions without sound, we can hypothesize that she might have *almost* reached her maximum performance level with more training. The sonification may has acted in this case as a catalyst for the smooth pursuit, allowing to reach a maximum level of performance more quickly. The subject shows in addition consistency in performance levels between the different patterns (see figure 6.13).

During the Trial phase, the opposite behavior is observable. The subject exhibits learning during the first part without sound (see figure 6.14, left), also with characteristic differences between the patterns. When sonification is used, the number of pursuits dramatically falls to a minimal level of 1.4 per session (over 20 trails), and stays this low for three consecutive sessions. Performance increases again in the last session, but only for selected patterns (the same as earlier in the experiment). These observations can be the sign of an important motor adaptation demanded by sound. The sudden perceptual changes led to the modification of still establishing sensorimotor plans, hence a decrease in performance ('de-learning' before 're-learning'). The number of saccades (right of figure 6.14) shows no important evolution and stays quite low, indicating few movements. It seems that the subject was perceptually affected by the sonification and needed time to initiate learning again (which is visible only in the last session). It is also possible that sonification provoked a high cognitive load in these difficult experiment conditions.
This subject illustrates that the effect of sound can be positive in one task (tracking) but also negative or perturbing in the free condition.

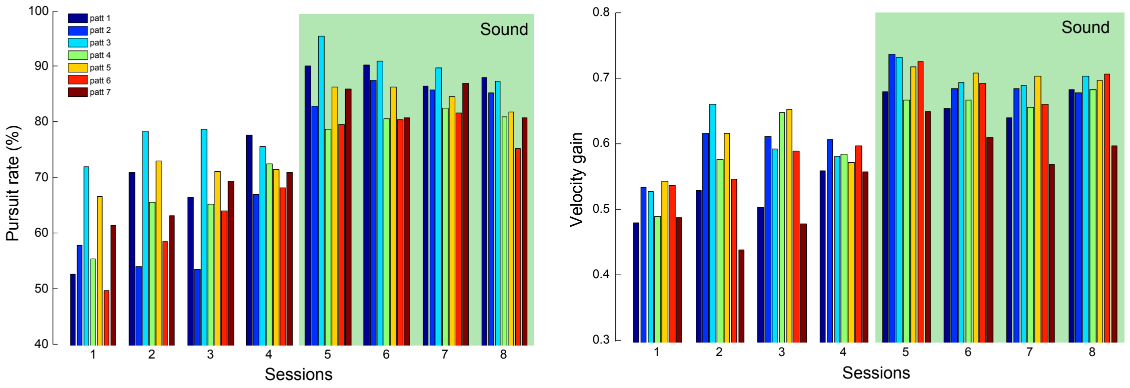


Figure 6.13 – Pursuit rate and velocity gain in Follow phase for subject #6.

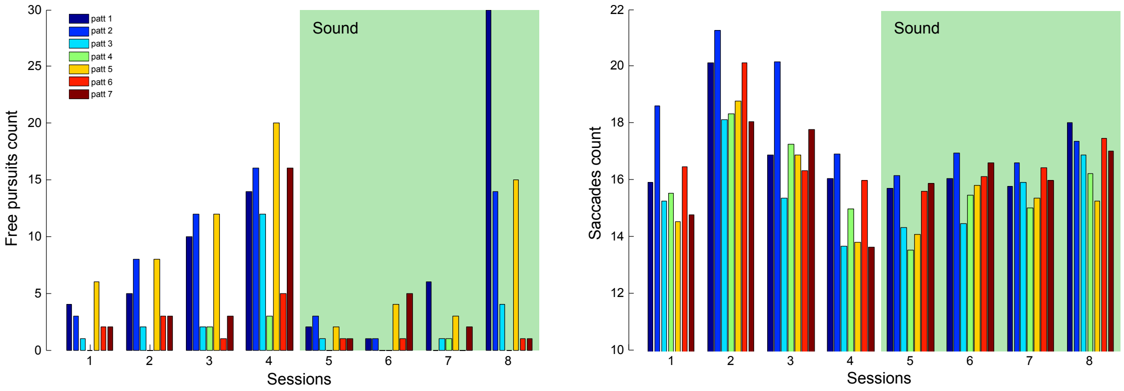


Figure 6.14 – Free pursuits count (sum) and saccades count in Trial phase for Subject #6.

Case of subjects #8 and 12 These two subjects exhibit similar behaviors, but are from different groups. In the Follow phase, pursuit rates and gain are significantly higher with the sonification, as can be seen figures 6.15 and 6.16, with a p-value of $p < 0.0001$ for both subjects. As for the previous case of subject #6, subject #8 exhibits an important performance increase when sonification appears. It can be noticed that this subject has mediocre pursuit rates (about 50%). Subject 12, who started the protocol with sonification, shows no sign of learning during the first four sessions (figure 6.16). After auditory feedback has been removed, the subject maintained his performance for two more sessions before they massively decreased (up to -40% for the pursuit rate).

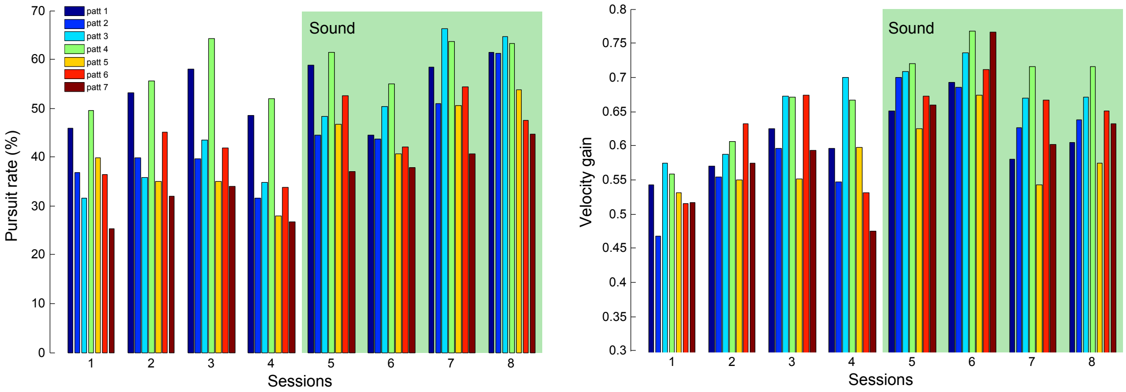


Figure 6.15 – Pursuit rate and velocity gain in Follow phase for subject #8.

In the Trial phase, both subjects also showed higher performance with sonification than without (left sides of figures 6.17 and 6.18). Statistical analysis revealed a significant increase ($p < 0.0001$) of the number of free pursuits. Subject #8 shows a fast learning curve in sessions 1 to 4, reaches maximum level in session 6 but hardly sustains his performances. The performances of subject #12 drops drastically when sonification is removed (especially since he produced many pursuits in sessions 3 and 4). In his case, learning with the auditory feedback seems to have induced a severe dependency on it, preventing consolidation effect to occur in the second part of the experiment.

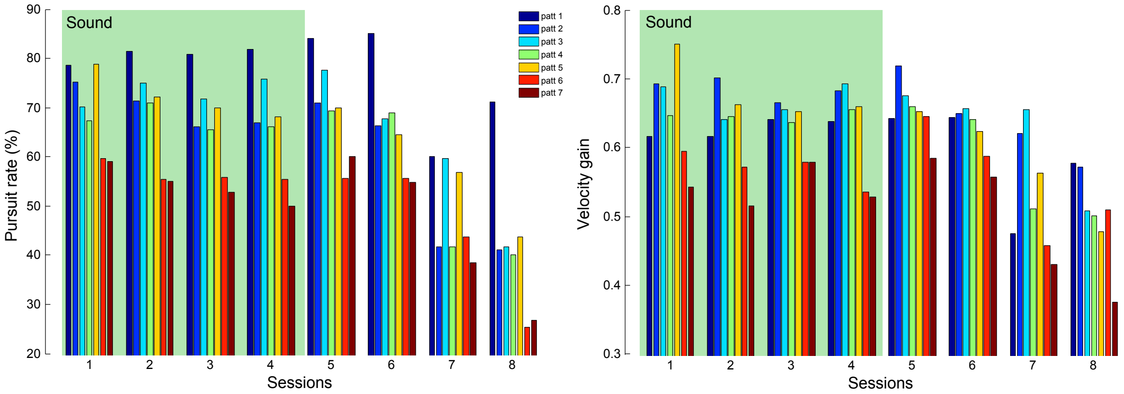


Figure 6.16 – Pursuit rate and velocity gain in Follow phase for subject #12.

For both subjects, the number of saccades in the Trial phase is correlated with the production of free pursuits (see figures 6.17 and 6.18). This suggests that sonification caused an increase of overall energy in their eye movements. This correlation could be used to characterize subjects who are the most responsive to movement sonification (*i.e.* more active).

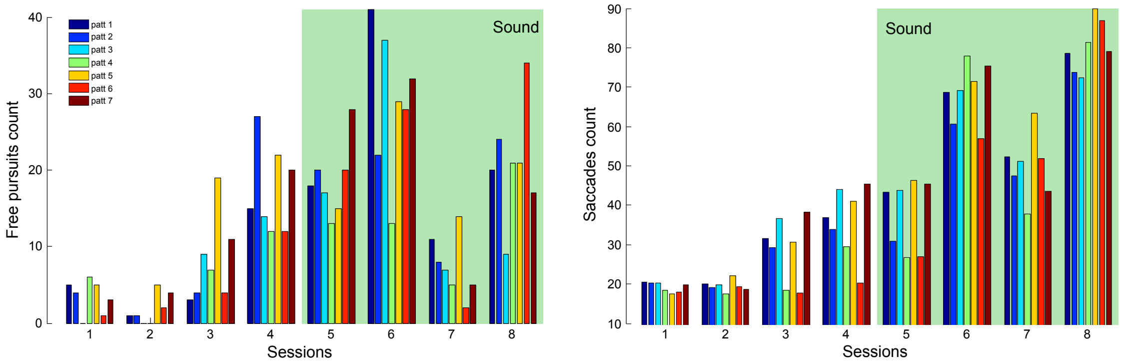


Figure 6.17 – Free pursuits count (sum) and saccades count in Trial phase for subject #8.

These observations tend to confirm the benefits of auditory feedback on instantaneous performance, but show also that performance is not robust to changes in the feedback conditions; *i.e* when sonification is removed, for subject #12 for instance. The guidance hypothesis, which characterizes a feedback-dependent learning (see section 2.2), could explain both antagonist behaviors that can be observed in the two groups. In the presence of sonification, the joint increase of tracking gain, free reverse-phi pursuits and saccades shows a positive reaction of the subjects to the sonification system.

Case of subject #10 (trained) This subject had already participated in an oculomotor experiment, which included target tracking, reverse-phi stimulus and free pursuits generation, one year before his participation. However, this experiment did not include any auditory feedback. His past training consisted in 10 sessions on one hour each (scattered over a month), during which he had to learn to produce numerical figures, letters, or

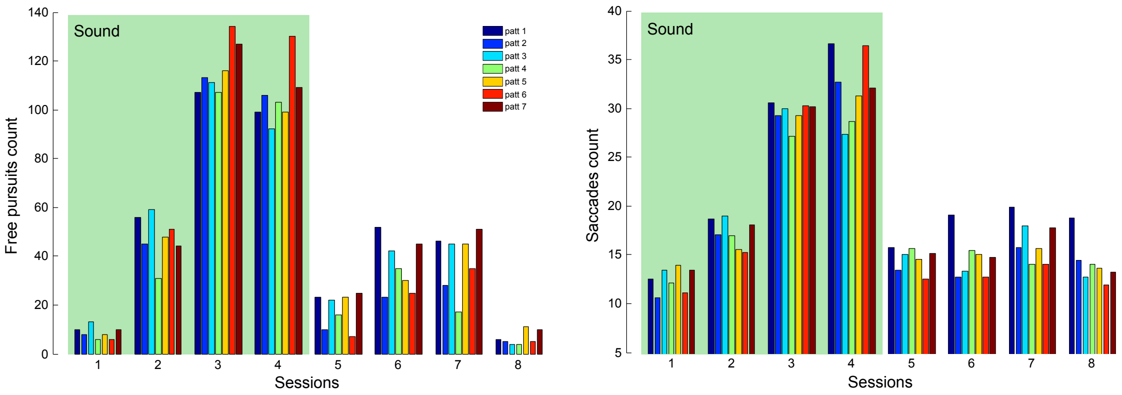


Figure 6.18 – Free pursuits count (sum) and saccades count in Trial phase for subject #12.

free shapes with a reverse-phi type stimulus. Preliminary target tracking exercises were performed during each session, although without the reverse-phi stimulus: back and forth horizontal and vertical movements and spiral drawing. The reverse-phi stimulus used was close to ours in terms of geometry (number and disks size) but contrast polarity was weaker, and flicker rate was 20 Hz, which is twice as fast as the one we used. We reiterate that this subjects was kept in the panel as an opportunity to observe his behavior as a non naive subject.

As expected, due to his previous training, the subject exhibited spontaneously excellent tracking performance in the Follow phase. His pursuit rate in the first session was 90%, with a gain of 0.85, one of the highest (see figure 6.19). Surprisingly, tracking performance actually deteriorated gradually session after session (about -20% in pursuit rate, -15% in gain at session 4). The arrival of auditory feedback in session 5 seems to slow somewhat the decrease and consolidate performance. It is actually not possible to rule out that this effect is only due to time. Statistical analysis showed a significant difference between sessions with and without feedback for both performance indicators ($p < 0.0001$).

Like for the Follow phase, his spontaneous performances in Trial phase are good (see figure 6.20). This illustrates the subject's understanding and ability to use the reverse-phi stimulus for free pursuits. As previously, pursuit production degrades gradually until session 4. In session 5, it drops again, and reaches a minimum at about -65% of the initial value. In the following sessions, the subject seems to learn how to produce free pursuits again (but does not even reach back the level of session 4). He particularly succeeds with patterns 3, 5 and 7. His performances are though significantly lower than in the Follow phase ($p < 0.0001$). The number of saccades produced remains quite stable, probably due to his experience in controlling them and of a reverse-phi stimulus.

The evident loss of performance for subject #10 raises important questions about the protocol we used and especially the revers-phi stimulus. Although no study[7] examined the influence of the reverse-phi parameters on free smooth eye pursuit movements, the higher

---

[7] Yet, and up to our knowledge.

contrast and slower flicker in our case may modify significantly the percept induced by the illusion (compared to the one this subject had been trained with).

Some expert users stand that complete learning of free smooth pursuit cannot be achieved with a strong contrast and a slow flicker. The idea is to limit the risks of stopping the eye if it is attracted by catchy events, like protruding points deeply flickering. As subject #10 has been trained with a quite different stimulus, especially faster (and for a longer time than the one he spent in this experiment), we suppose that the surprising results we observed are due to a strong adaptation required to the 'new' reverse-phi stimulus. The arrival of the auditory feedback in the middle of the training have probably caused another need for sensorimotor adaptation, whereas the subject was already 'de-learning' and trying to adapt. The stabilization of his performances by the end of the experiment seems coherent with this interpretation.

Furthermore, this gradual degradation also appeared in the Follow phase. Although smooth pursuit of a moving target is an intrinsic motor ability for humans, (and the fact that this subjects performed well at the beginning), this task has been proposed here with the reverse-phi stimulus in the background, which could have perturbed the subject in this phase as well. We hypothesized that in the absence of reverse-phi stimulus, his performance would not have been greatly affected during the Follow phase.

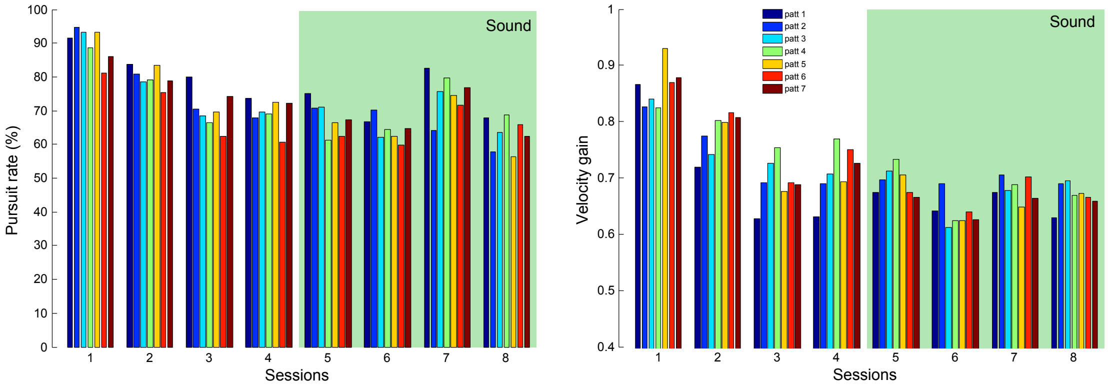


Figure 6.19 – Pursuit rate and velocity gain in Follow phase for subject #10.

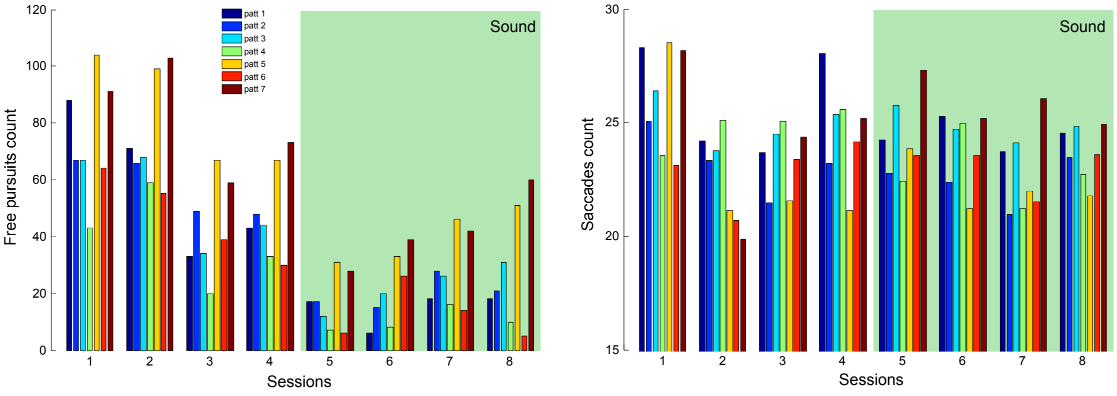


Figure 6.20 – Free pursuits count (sum) and saccades count in Trial phase for subject #10.

Questionnaires The visual fatigue was rated 2.3 ± 0.5 over 10 by the subjects. On 96 total ratings, 18 of them were 5 or more, specifically for three particular subjects. The visual fatigue experienced is thus moderate. No particular auditory fatigue has been reported (0.25 on average, no more than 3/10). To the question "Do you think you became aware of your gaze position since the beginning of the training?", subjects answered 4.1 ± 1.8 over 10, with a maximum in the sessions of 5.7 ± 2.1, meaning they did only slightly. Ratings about a feeling of better gaze control after training started were 3.4 ± 1.5 on average and maximum rating was 5.1 ± 1.8. Before the experiment, few subjects declared to pay attention to their gaze; this concerned only subjects who practice fast action sports (tennis) or serious readers.

Subjects rated 3.9 ± 0.3 over 10 the affirmation that sound helped them accomplishing the task (free pursuits). On the 48 total ratings, 19 were 5 or more. It remains delicate to interpret this value, which depends on each subject's self judgment of success. Nevertheless, none of the subjects seemed indifferent to auditory feedback, neither according to their ratings, nor judging by their spontaneous reactions and comments.

In self evaluation of performance, each subject established his own criteria for assessing his improvements and absolute level of performance. This caused large variability in the ratings. It appears that these ratings are not correlated with their actual performance level. Judging by the actual productions, we suppose that subjects rated more the global shape of the pattern they tried to draw than the quantity or quality of the smooth pursuits they had to produce. This global image is all the more difficult to evaluate for them that they never received knowledge of result about the spatial trajectory they produced.

Nevertheless, a strong tendency can be observed. Group 1 subjects (who started without sonification) exhibited a large drop of their ratings after the first session with sound. Subjects strongly depreciated their personal performance, regardless of the previous ratings in the past four sessions or their evolution. No clear behavior points out of group 2. This could indicate that self evaluation is all the more precise that it is performed during sessions with auditory feedback; subjects may have calibrated their ratings with the help of the sound, and could have been less able to adjust them without auditory feedback. Group 1 subjects did correct their impressions when sonification was present, indicating a stronger confidence given to auditory information about performance. This observation confirms that the subjects were well aware of how to interpret the auditory feedback.

6.4 Discussion

The study reported in this chapter presents an unexplored use of sonification by targeting learning of specific ocular movements. We faced two main difficulties in this study: the large inter-individual variability and the difficulty of subjects to perceive and acquire the reverse-phi illusion.

We found that good tracking results were not correlated with the ability to learn to perform free smooth pursuits. This can be a consequence of the difficulty to perceive and control the reverse-phi illusion, which can be independent from tracking abilities. Similarly, no direct link was found between free pursuits production and the quantity of saccades: more saccades could mean either more energetic behavior or less control. The variety of behaviors observed emphasizes that more specific analysis methods should be developed for identification and modeling of tracking and free smooth pursuit performances.

Eye movement sonification had a positive effect on six subjects in the Follow phase, improving duration and/or precision of the tracking movements. Moreover, it has not been observed that sonification could degrade the performance in this task. This could be explained by a preponderance of visual information: if auditory information are not necessary, they may be ignored in favor of visual ones[8].

In the case of free pursuits production, ten subjects did react to the appearance (or disappearance) of sonification during training. For six of them (50% of the panel), sound had a positive effect on performance. In the free condition, auditory information seemed to overcome visual percept, mainly because it indicates the appearance of smooth pursuits using another sensory channel. The same observation was found with the self evaluation ratings. This "decoupling" of the action (the eye motion) and perception cannot be complete though, as the very perception of the visual illusion is a necessary and sufficient condition for initiating pursuits.

An important number of subjects exhibited what we interpreted as negative reactions to the auditory feedback, especially saccades sonification. As saccade production is the main parameter that needs to be controlled in the free condition, an excessive sonification can be perceived as a negative reward and restrain learning. Also, although sound feedback reflects the eye behavior, a lack of 'aesthetics' have been pointed, saccades being perceived as "gun shots". The question of the 'affective' design seems sensitive in the case of eye movements sonification (also because these experimental conditions can be tiring).
Additionally, the complexity of the task proposed, and the 'unusual' or 'disturbing' aspects of the perception-action loop, require first the establishment of explicit learning to achieve pursuit production. The appearance of auditory feedback in the middle of the learning process (for group 1) could provoke sensorimotor adaptation that stops or impairs the initial learning. This could explain the sudden drop in performance that some subjects exhibited. Additional time is then needed to integrate the novel sensory conditions.

The example of subject #12 illustrates the issue of persistence of learning and a possible dependency on feedback (named "guidance hypothesis" (Buchanan and Wang, 2012; Ronsse et al., 2011)). Although this subject benefits from sonification, he is unable to build a robust learning. As mentioned in chapter 5, the design of the experimental protocol and auditory feedback should take into account this possible effect to ensure robust and long

---

[8] A sensory Ockham's razor?

term learning. The results presented, especially the differences between the two groups, indicate that this particular task requires long training periods, and that pre and post tests should be carried out to evaluate feedback dependency.

This study also shows the suspected influence of parameters of the reverse-phi stimulus on free pursuits learning. The flicker rate and contrast range may affect strongly the possible dynamics of free pursuits. The example of the previously trained subject illustrates this point. His performance results indicate that he could not maintain high pursuit production under these new conditions, he then had to adapt and re-learn the task. The speed of 'de-learning' processes have been noticed by Davidson and Wolpert (Davidson and Wolpert, 2004), who indicated that motor learning a new movement may require hundreds of repetitions, whereas the inverse mechanism was much faster.

One can discuss whether auditory feedback could be classified as knowledge of result or knowledge of performance in this case, and whether such a classification is pertinent. Auditory feedback provides information at different time levels, and continuously. Firstly, auditory information has to be integrated over a whole trial to get an image of the performance (the amount of smooth pursuits). Secondly, saccade information is delivered instantaneously; the subject gets an acoustical image of the saccade allowing him to rapidly associate this information with the visual outcome of his action. In addition, the pursuit sound provides an auditory image of the visual illusion created by the stimulus and the motion, but does not *directly* indicate how to initiate free pursuits. In the case of continuous sonification of time-varying variables, the limits of KP and KR notions are sometimes reached. This formalism issue is also addressed in chapter 7.

The study suggests that achieving control of free pursuits was subjected to a prior perceptual learning of the reverse-phi illusion (Darainy et al., 2013; Fine and Jacobs, 2002). The key element being the ability to keep perceiving the illusion while initiating pursuits. The great difficulties encountered by the subjects show that a first phase of strong familiarization with the context is needed. A sort of 'eureka' effect has been observed in some subjects, when they suddenly succeeded the task, indicating that this familiarization time varies largely between individuals. The individualization of the learning protocol is an important path to address for this type of task. The use of sonification can be a good solution, which allows to emphasize visual and proprioceptive feedback. The aim being to acquire the targeted perceptual state, taking advantage of sound to deliver low level and non ambiguous information.

6.5 Perspectives

Although our results are preliminary, we observed that sonification of eye movements can have an effect (positive or negative) on different aspects of track and free pursuit performance. Even if our results cannot be generalized, this exploratory study raises intriguing questions, which motivates future works.

The ‘oculo-auditory-motor’ paradigm we explored in this study represents a new path for investigating auditory-motor coupling in a very specific system. We precise that most of the subjects could hardly believe (quoting their own words) that the sounds the were hearing were actually caused by their movements. The sensations felt by hearing this audio energy, synchronized with the weak proprioception of eye movements are quite fascinating. We plan to have experts of oculomotor control experience our system to get their feedback. An important part of the future work will be designing enjoyable sound textures, especially regarding the sonification of saccades energy profile.

It is necessary to extend our knowledge of pursuit movements with reverse-phi stimulus in order to optimize the sonification strategies. It is also unclear whether combining tracking exercises and free pursuit can be beneficial. It seems that control strategies will vary according to the targeted pattern and associated difficulty (as it is observable in this study). The consistency of the relative success for the different patterns confirms this idea.

Much further research is needed to investigate the use of continuous sonification of oculomotor learning. Using well known protocols of the vision field, such as target occlusion, or pursuit-induced movement in the dark, and bringing sonification to them can be a good strategy, which we are currently considering. Questions could be whether reverse-phi stimulus allow for maintaining tracking during occlusion of a moving target, or whether continuous sonification of eye dynamics can do so too.

Similarly to the observations made in chapter 5 for a visuo-manual tracking task, we propose to further explore the particular sensorimotor loop of sonifying intrinsic movement characteristics (and not the error relatively to the task). Studies on ocular proprioception like (Steinbach, 1987; Tong et al., 2008) could benefit from auditory feedback designed to augment proprioceptive feedback.

Specific applications can be reasonably evoked, based on fundamental and creative aspects of our approach. Both research and industrial applications can be targeted. This work opens towards other interdisciplinary research such as HCI, cognition and perception, but also interaction design. Systems like the one we developed can be extended to other features of eye movements, taking advantage of the energy of saccades, gaze stability control or pupil diameter. We could imagine pedagogical applications of sonification for developmental experiments and for orthoptics practice.

Last but not least, communication interfaces or musical controllers could take advantage of this approach. As we pointed out earlier, digital musical interfaces could gain expressiveness if the continuous dynamics of eye movements is integrated as a control dimension. But one of the most important application being targeted is the development of communication interfaces for paralyzed patients, who could benefit from eye movement sonification for training how to write with their eyes, and gain engagement or enjoy artistic activities.

Chapter 7

Learning movement kinematics with a targeted sound

E. O. Boyer[1,2], Q. Pyanet[1], S. Hanneton[2], F. Bevilacqua[1]
[1] IRCAM, STMS-CNRS-UPMC, Paris, France
[2] LPP, Paris Descartes CNRS UMR 8242, Paris, France

Refereed article published in *Sound, Music, and Motion*, Springer International Publishing, 218-233, 2014.

Abstract: This study introduces an experiment designed to analyze the sensorimotor adaptation to a motion-based sound synthesis system. We investigated a *sound-oriented* learning task, namely to reproduce a targeted sound. The motion of a small handhold object was used to control a sound synthesizer. The object angular velocity was measured by a gyroscope and transmitted in real time wirelessly to the sound system. The targeted sound was reached when the motion matched a given reference angular velocity profile with a given accuracy. An incorrect velocity profile produced either a noisier sound or a sound with a louder high harmonic, depending on the sign of the velocity error. The results showed that the participants were generally able to learn to reproduce sounds very close to the targeted sound. A corresponding motor adaptation was also found to occur, at various degrees, in most of the participants when the profile is altered.

Keywords: Gesture, Sound, Sensorimotor, Learning, Adaptation, Interactive systems, Auditory feedback, Sound-oriented task.

7.1 Introduction

There is growing interest in using tangible interfaces and motion sensing technology to interact gesturally with digital sound processes. In particular, a research community has been established over the last ten years around the development of gestural digital musical instruments (DMIs). The NIME conference (New Interfaces for Musical Expression) (Bevilacqua et al., 2013a) has centralized several important research results. While the evaluation methodology of such interfaces is recognized as important, it has generally been considered from a user experience point of view, most often ignoring fundamental aspects of sensorimotor learning. Nevertheless, we believe that sensorimotor learning should be fully addressed for the development and evaluation of digital musical interfaces.

This research topic is close to applications using *movement sonification*, where digital sound processes are designed to react to movements, hence providing additional information about the movement/performance. Typically, the auditory feedback is thought to supplement other sensory modalities (such as proprioception and vision) and to facilitate sensorimotor learning. Such an approach has been proposed for example for the facilitation of skills acquisition in sports (Wolf et al., 2011) or in physical rehabilitation (Robertson et al., 2009). Although there is a growing number of publications studying the mechanisms whereby auditory feedback can improve motor control and learning, there is still a lack of formalism and consensus on the use of such auditory feedback.

We have started to study sensorimotor learning in DMIs and interactive sound systems for movement training/rehabilitation, within a single research project[1]. We take advantage of the fact that these applications can share identical technology (motion sensing and processing) and also share similar questions about the action-perception loop involved in motion-sound interaction.

While the different applications might imply similar sensorimotor learning processes, they can still be categorized based on the different tasks they imply. In the case of DMIs, the task can be expressed as *sound-oriented*. The users adjust their movements in order to achieve a specific goal expressed in terms of sonic/musical characteristics. In the case of motion training (i.e. sport or rehabilitation), the task can be expressed as *motion-oriented*. The users get auditory feedback to adjust their movements and to achieve a specific goal in terms of motion characteristics. In Figure 7.1 we schematically describe the information flow of movement sonification in the cases of *motion-oriented* and *sound-oriented* tasks. The figure emphasizes that both concepts share the same architecture.

In this paper, we focus only on the first case: the *sound-oriented* task. In this case, the user's attention is drawn on the sound produced by the action and the auditory-motor loop is regulated by the perceived sound properties. We present an experiment where we evaluate the movement adaptation of subjects who are asked to control a specific sound quality.

The paper is structured as follows. First, we present a short overview of related works. Second, we describe the experimental setup, methodology and motion analysis. Third, we present the results, and fourth, we discuss our findings and their implications for further experiments.

---

[1] Legos project, http://legos.ircam.fr

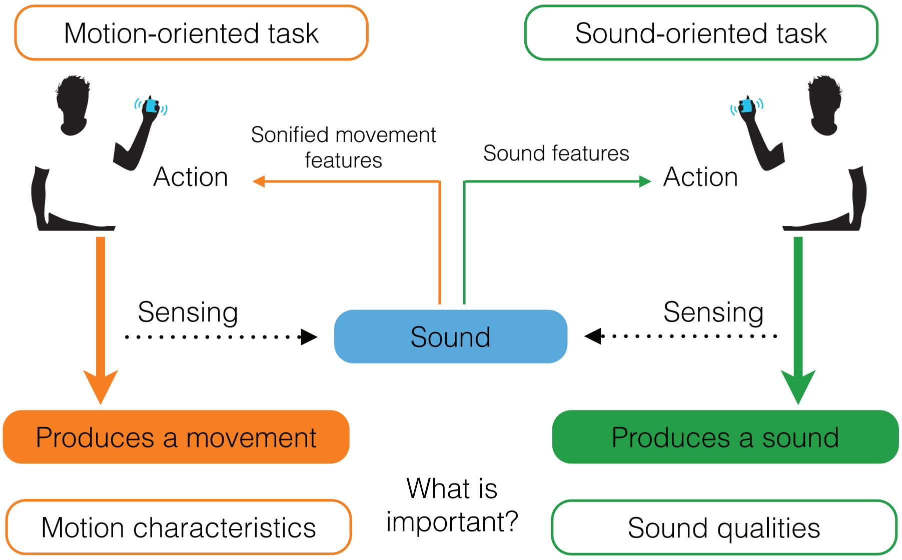


Figure 7.1 – Concept of Motion-oriented vs Sound-oriented task.

7.2 Related Works

We first describe here the few studies that explicitly reported on a *sound-oriented* task. We then report on other *motion-oriented* tasks that showed the interest of using auditory feedback.

7.2.1 Sound-oriented Task

A small number of studies have examined the concept of *sound-oriented* task. Early works were performed focusing on the evaluation of gesture-sound mappings. Hunt et al. in (Hunt and Kirk, 2000) presented such an evaluation by asking subjects to reproduce a target sound using different mapping strategies. Only simple interfaces such as a mouse and sliders were used. It resulted that, while complex gesture-sound mappings were more difficult to master, they appeared to be more engaging for the subjects. This implies that the type of implicit learning involved in this case was perceived as beneficial.

Gelineck et al. (Gelineck and Serafin, 2009) also studied input interfaces and compared knobs and sliders for a task consisting in reproducing reference sound samples. Subjects were musicians and were asked to reproduce four sounds with temporal timber variations (synthesized with a physical model of flute and friction). A qualitative evaluation was performed showing that no significant difference was found between the use of knobs and the use of sliders. Note that these studies did not explicitly address sensorimotor learning or adaptation in their questionnaire-based evaluation.

Pointing towards auditory targets can also be considered as a sound-oriented task. Recently, we investigated the effect of sound feedback on blindfolded pointing movements towards auditory targets spatialized with HRTF binaural technique (Boyer et al., 2013).

We found that the auditory target should last enough to be heard during the task. The potential advantage to additionally sonifying the hand was not apparent in such a case. Forma et al. (Forma et al., 2011) showed that blindfolded participants are also able to point towards targets in a auditory virtual environment using sound as sensory substitution. Interestingly participants succeeded in the task whether the virtual listener was congruent with their ears or placed on one of their hands.

The concept of a *sound-oriented* task can be linked to recent studies on the relationship between body motion occurring during various sound/music stimuli (Godøy et al., 2006b; Godøy et al., 2006a; Caramiaux, 2011; Nymoen et al., 2010; Nymoen et al., 2011). In particular, Godøy et al. (Godøy et al., 2006a) investigated motion trace that subjects performed on a 2-dimensional surface in response to a sound stimuli. Other studies were reported on hand gestures performed while listening to either abstract synthesized sounds (Nymoen et al., 2011), or stimuli derived from environmental sounds (Caramiaux et al., 2014). As expected these studies showed that the motion related to sound stimuli depends on several different sound aspects and varies greatly between subjects. Nevertheless, such studies offer novel perspectives in showing experimentally that some sounds can favor specific motions.

7.2.2 Auditory Feedback in Motion-oriented Task

The other types of related studies concern investigations of *motion-oriented* tasks to establish whether auditory feedback can be beneficial for learning and performance. Rath and Schleicher (Rath and Schleicher, 2008) studied a virtual balancing task under different feedback conditions, including auditory feedback to guide movements. They found that the auditory feedback was beneficial in terms of rapidity, the best results being found when sonifying the ball velocity. They also found small differences between ecological and abstract sounds. More recently, Rosati et al. (Rosati et al., 2012) showed that a tracking task can be improved using an auditory feedback (in addition to a visual feedback) related to the task achievement or, to a lesser extent, giving information about the error.

Vogt et al. (Vogt et al., 2009) proposed a movement sonification system to improve perception of body movements. Sonification and “positive”sounds were beneficial for task understanding and increased the subject motivation. Effenberg (Effenberg, 2004) focused on an ecological approach, insisting there is a close relationship in kinesiology between movement kinetics and sound. He showed that supplementary auditory information improves the perception and reproduction of sport movements compared to vision alone. These results appeared independent from the qualitative assessment of the sounds qualities by the subjects. Takeuchi (Takeuchi, 1993) previously pointed out that sound is a very useful information channel in sports. Avanzini et al. (Avanzini et al., 2004) insist on the role played by auditory information in multi-modal interactions. Wolf et al. (Wolf et al., 2011) and Effenberg et al. (Effenberg et al., 2011) showed that subjects can benefit from multi-modal motor representation in a rowing-type task.

Wolf et al. also report that auditory feedback can reduce spatial error and improve synchronization when the feedback is related to the internal representation of the task rather than short-time features of the movement. Karageorghis and Terry (Karageorghis and Terry, 1997) suggested as well that sound feedback can improve mood, hence performance, in sports and leisure activities.

Sport and musical control are not the only domains where auditory interaction can improve motor learning. Thoret et al. (Thoret et al., 2012) studied the sonification of drawings to investigate whether subjects could recognize a drawn shape from recorded and synthesized friction sounds. They noticed that people were able to identify gesture trajectories with the friction sound they produced and the model-generated sounds which used movement velocity as input.

A recent review by Sigrist et al. (Sigrist et al., 2013) presents experimental studies of sonification techniques. The authors formalize the different type of auditory feedback in the framework of motor learning theories. They insist on the fact that, despite important applications, several questions on auditory feedback for motor learning remain insufficiently explored.

Recent studies show that an additional feedback can improve physical rehabilitation processes and there is growing interest in using additional auditory feedback to guide movements of impaired or stroke patients (Robertson et al., 2009; Subramanian et al., 2010; Avanzini et al., 2009; Rosati et al., 2011). For instance Huang et al. (Huang et al., 2005) designed a multi-modal biofeedback with musical tracks in a reaching task with stroke patients and found that visual and auditory feedback together helped patients producing smoother and more accurate movements.

7.3 Materials and Methods

7.3.1 Experimental Setup

The *sound-oriented* task is based on the manipulation of a specific motion interface that allows for the continuous control of sound synthesis. Subjects are seated in front of a table on which two spots are drawn, named 1 and 2, marking the spatial starting and ending areas of the movement they will have to make. Subjects carry in their hand the motion interface, consisting of a small object containing 3D accelerometers and a 3-axis gyroscope. Figure 7.2 depicts schematically the setup. Data are transmitted wirelessly to a receiver through the IEEE protocol 182.15.4 (2.4 GHz Band), that transmits them to the computer using Open Sound Control (through the UDP protocol). A software programmed using the Max/MSP environment[2] includes real-time data processing, sound synthesis and data

---

[2] `www.cycling74.com`

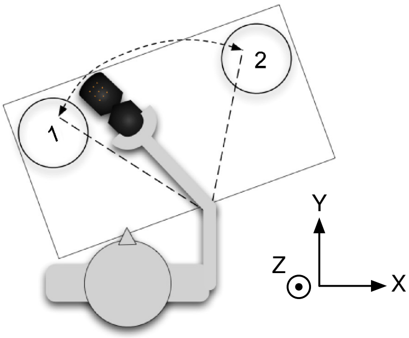


Figure 7.2 – Experimental setup. The subjects move the tangible interface from 1 to 2 in order to continuously control the sound and aim the targeted sound.

logging (data, sound and video recordings of each subject). The subjects listen to the sound using headphones.

The angular velocity around the Z axis of the interface is used as input. The sound is synthesized from the difference between the performed velocity profile and a defined velocity profile, the *reference profile*, that varies between different conditions. This profile is a bell shape curve (derived from a Gaussian profile), corresponding roughly to the velocity profile typically found while moving the hand between two points (Mitrovic et al., 2010), with a maximum peak velocity around 70 deg.s^{-1}.

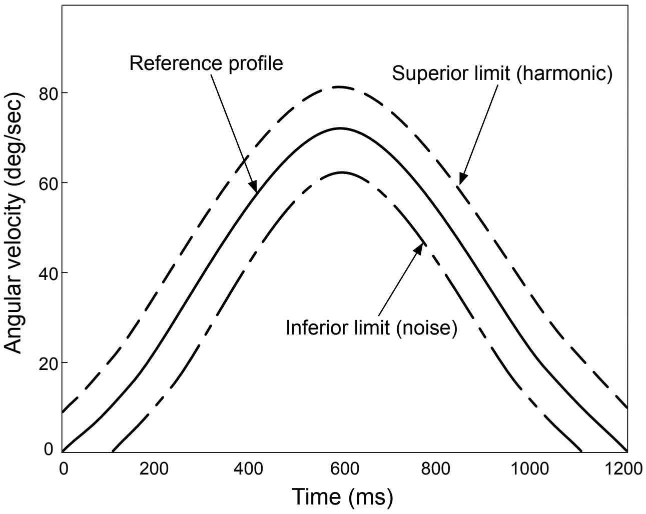


Figure 7.3 – Reference profile and the associated thresholds enabling the change in the sound qualities (noise or loud higher harmonic).

The velocity signal is mapped to a sound synthesizer using Modalys[3] in Max. A resonator, a string model, is used to filter three types of input sound signal: one square sound signal at a fundamental frequency equal to 260 Hz (corresponding to C4), matching

---

[3] Modalys (Ircam), `http://www.forumnet.ircam.fr/product/modalys`. The object used is “MONO-STRING”, see documentation for details.

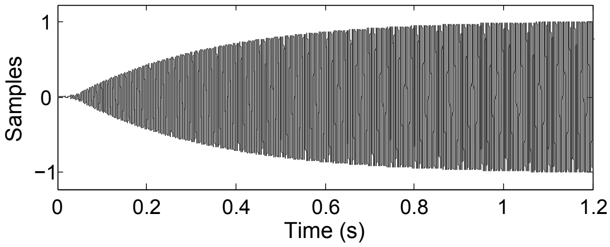


Figure 7.4 – Waveform of the target sound.

the second harmonic of the string, one square sound signal at a fundamental frequency equal to 910 Hz, matching approximately the 7th harmonic and pink noise (constant power per octave). The difference between the performed profile and the reference profile modulates the intensity of the higher harmonic or the noise inputs : positive values boost the higher harmonic, negative values boost the noise sound. This choice is motivated by the analogy with the velocity/pressure adjustments in bowing a string in a violin: low velocity might produce a noisy sound (with sufficiently high pressure), while increasing the velocity produces higher frequencies.

The sound level of the added effect is effective only when the difference reaches a given threshold, of constant value over the whole profile of velocity, as illustrated in Figure 7.3. Once the threshold is reached, the intensity of the effect depends linearly on the difference between the performed and reference velocity values. Our interest was to investigate how the subjects can learn a specific movement without guessing it directly from the sound morphology of the reference sound. As shown in Figure 7.4, the intensity morphology of the target sound does not match the reference profile.

7.3.2 Experimental Procedure

The subjects first listen to the target sound and to typical sounds associated to incorrect movements: one with noise referring to a lower angular velocity movement and one with an extra harmonic note referring to a higher angular velocity. All the sounds are 1.2 seconds long. The subjects can listen to the sounds as many times as they wish until they feel comfortable distinguishing the different sound characteristics. Figure 7.5 shows the spectrogram of the three example sounds, which are chosen to be easily discriminated by the subjects according to their frequency content.

Subjects are then instructed to move the object with their dominant hand between areas 1 and 2 to produce the target sound. Their motion should last as long as the sound (1.2 s). The subjects do not have control anymore on the sound produced beyond 1.2 seconds.

During the first phase, we call *Exploration*, subjects perform 60 movements (30 rightward and 30 leftward) with the normal symmetrical profile called E as a reference for feedback

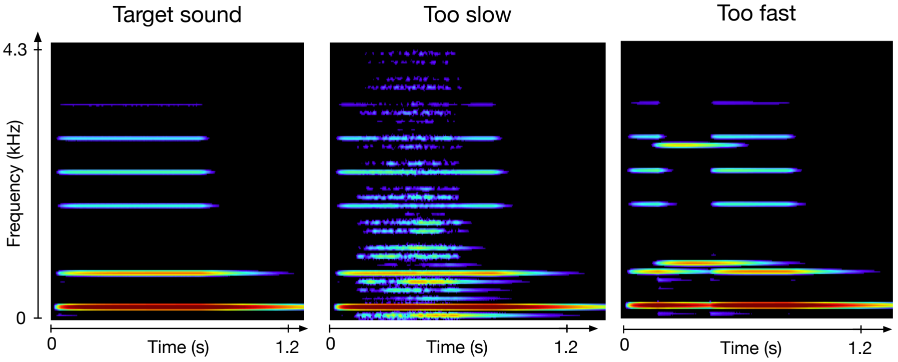


Figure 7.5 – Sonograms of the three example sounds; left: target sound produced if the movement matches exactly the target velocity profile *i.e.* the plain line in figure 7.3; center: sound containing noise produced by a slow movement; right: sound with the higher harmonic produced by a too fast movement.

generation. Between each movement, they must wait until a *beep* is emitted, which occurs randomly between 2.5 and 3.5 seconds. This random start is set to avoid the creation of a rhythmic pattern in chaining the movements.

In the second phase, *Adaptation*, subjects are blindfolded and asked to perform three blocks of 50 movements. For each block, the reference velocity profile was changed following the sequence A - B - A, without informing the subjects. As illustrated in Figure 7.6, the profiles A and B were obtained from profile E by shifting the temporal position of the maximum velocity. Figure 7.6 also shows the changes in the initial slope, which is approximated between start point and maximum of the profile. Profile A thus exhibits a higher acceleration and a slower deceleration. Profile B exhibits the opposite variations: a lower acceleration and higher deceleration.

The subjects are asked to fill in a questionnaire at the end of the experiment. It contains questions about their musical abilities, whether they are used to manipulate digital interfaces, asks whether they noticed modifications in the system in both phases, and invites them to rate the difficulty and the degree of control they experienced.

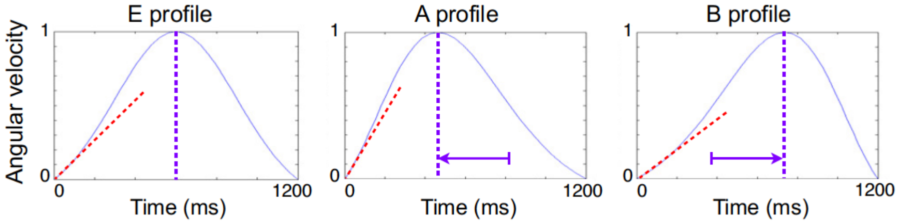


Figure 7.6 – Reference profiles of angular velocity used in the different phases of the experiment, showing initial slopes and peak shifts; amplitudes are normalized.

7.3.3 Subjects

Fourteen subjects volunteered for the experiment. All were healthy and reported normal hearing. They were 23.6 ± 1.9 years old and three of them were left-handed (21%). All were familiar with digital interfaces such as computers, and were familiar with music from recreational to professional levels (1 to 20 years of instrumental practice). All subjects gave written informed consent for the experiment.

7.3.4 Data Analysis

The analysis is based on the comparison between the angular velocity time profile performed by the subjects v_i and the reference profile u_i, where i is the i^{th} time sample (50 Hz sampling frequency). The recorded profiles are low-pass filtered with a 10 Hz cutoff Savitsky-Golay filter. As described below, different measures are estimated to capture specific features of the performed profiles. In a second step, the time evolutions of these measures were examined to find trends over the series of the subjects' trials, using t-tests and ANOVAs.

7.3.5 Angular Velocity Profile Parameters

The different measures described below were considered:

First, the mean error can be evaluated for each trial by taking the standard deviation of the difference between performed angular velocity v_i and the reference profile u_i:

$$mean\ error = \frac{1}{(N-1)} \sqrt{\sum_{i=1}^{N} [v_i - u_i]^2} \tag{7.1}$$

N being the total number of samples.

Second, the mean or first order moment of the profile was computed. It allows us to characterize where the largest velocity values are reached.

$$first\ moment = \Delta t \frac{\sum_{i=1}^{N} v_i i}{\sum_{i=1}^{N} v_i} \tag{7.2}$$

Δt being the time interval between two samples.

Third, we computed an approximation of the initial slope of the velocity profile. The changes in dynamics to A and B profiles come from the modification of the last two parameters, it is thus natural to compute these measures to evaluate subjects adaptation to the changes in the mapping. Table 7.1 gathers the parameter modifications in profiles

E, A and B.

Table 7.1 – 1^{st} order moment and initial slope of the different reference angular velocity profile phases.

Profil	1^{st} moment [ms]	initial slope [deg]
E	600	34.6
A	536	41.0
B	684	28.4

7.4 Results

We first investigated the evolution of the performance by comparing average error values at the beginning (8 first movements) and at the end (8 last movements) of each block (E, A, B, A). A general statistical analysis (ANOVA) was performed with three factors: the 4-level 'block' factor, the 2-level 'beginning/end' factor and the 8-level 'trial' factor. The analysis revealed a significant effect of the 'beginning/end' factor alone ($F_{(1,13)}$=26.3, $p<0.005$) which was not the case for the 'trial' factor. The interaction of 'beginning/end' and 'block' factors interestingly presented a significant effect on the performance ($F_{(3,39)}$=9.2, $p<0.005$), but the post-hoc tests indicated significant error reduction only within the first block (the Exploration phase). This shows that there is significant learning occurring in the Exploration phase which we further examined using individual t-tests.

7.4.1 Exploration Phase

During the Exploration phase, each subject starts with a spontaneous motion from area 1 to 2. By listening to the auditory feedback, they are able to adapt their movement to reach, more or less, the reference profile. A typical example is shown in Figure 7.7, where the first and last profiles are plotted along with the reference profile. In this case, the ending profile is clearly closer to the reference profile than the initial one.

The mean error values of the velocity profile are shown in Figure 7.8 for each subject. Error bars indicate the standard deviation across the profiles for a given subject. A large variability between the subjects can be observed on the initial movements (dark gray bars). This was expected since no specific instruction was given to the subjects about the dynamics of the movement they had to perform. These differences can thus be directly linked to the variability of the spontaneous movements performed by the subjects. After more than 45 trials, the variability between the subjects is largely reduced (by 50%), which indicates they were able to use the sound feedback to constraining their motion towards the reference profile.

Importantly, Figure 7.8 also shows that for all subjects the mean error is lower in the last trials than in the first trials, which is also a strong indication of the positive effect of the

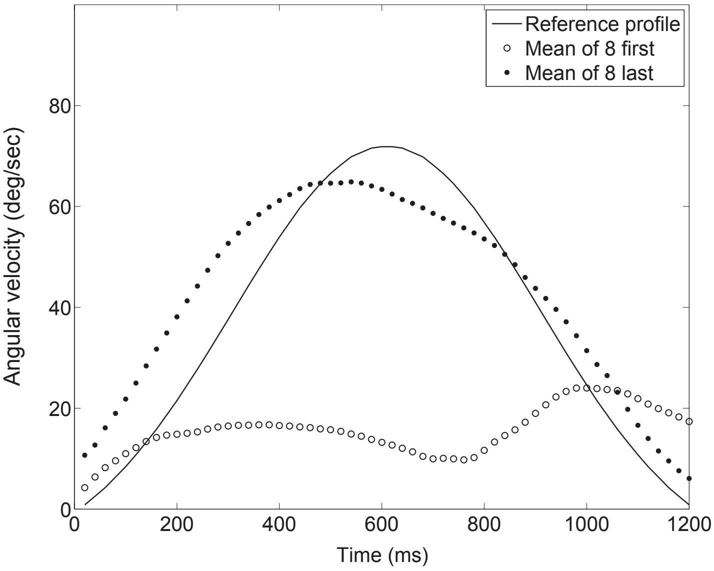


Figure 7.7 – Example of angular velocity profiles during the exploration phase (subject #7). The comparison between the first and last profiles clearly shows that the subject modified his movement towards the reference profile.

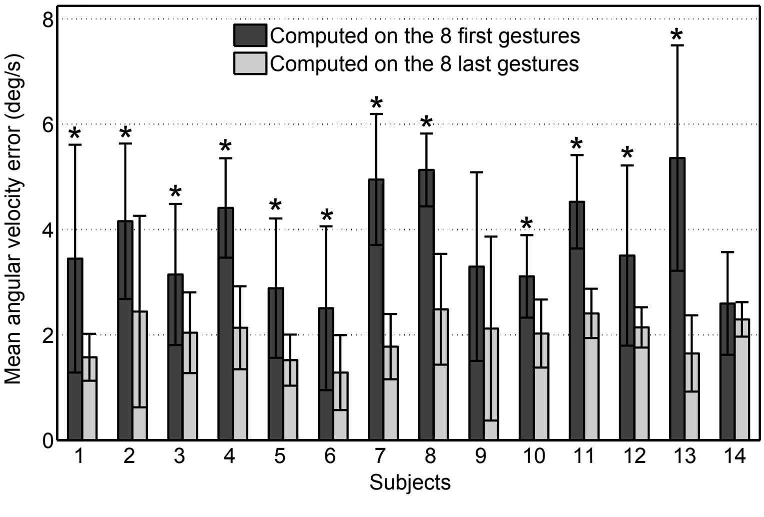


Figure 7.8 – Mean error results on angular velocity profile for each subject during the Exploration phase E; error bars indicate standard deviation; the asterisks indicate significant error reduction at the end of the Exploration phase ($p \leq 0.05$).

auditory feedback. To characterize this quantitatively, we performed *t*-tests to determine which subjects exhibited statistically significant improvements ($p < 0.05$ shown with an asterisk in Figure 7.8). This result confirms the general ANOVA performed previously, and provides us with more detailed information: 12 subjects out of 14 significantly adapted their motion during phase E. Among the two subjects who did not show significant improvement, subject #14 spontaneously performed motions with errors relatively close to that of last profiles for the other subjects, which might explain why the improvement was less significant. Subject #9 exhibited large standard deviations which also explains why the improvement is not statistically significant. The Adaptation phase discussed in the next section provides more information about the performance of these subjects.

7.4.2 Adaptation Phase

During the Adaptation phase, the A and B profiles are alternated, which allows for a more detailed investigation of the subject performances. We emphasize that the subjects were not informed of the change between the A and B profiles. The main difference between these profiles can be characterized by the variations of the first moment, or by the initial slopes (see Table 7.1). The first moment is actually close to the relative time to peak velocity (rTPV). Nevertheless, we found the computation of rTPV less robust, due to irregularities sometimes occurring in the velocity profiles. Therefore, we focused on the first moment and the initial slopes and performed statistical tests to examine whether significant adaptation can be observed within the transitions A to B and B to A. The results are reported in Table 7.2.

Table 7.2 – Significance of the parameter variations during the Adaptation phase, between the 14 last trials of each block ($p \leq 0.05$)

Subject	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1^{st} moment A → B	⋆	⋆		⋆		⋆	⋆		⋆	⋆	⋆		⋆	⋆
1^{st} moment B → A	⋆	⋆	⋆			⋆	⋆				⋆		⋆	
initial slope A → B		⋆		⋆		⋆	⋆			⋆	⋆	⋆	⋆	⋆
initial slope B → A		⋆	⋆	⋆		⋆	⋆			⋆	⋆	⋆	⋆	

The individual *t*-test results show that we can separate the subjects into three groups. First, 5 subjects show significant adaptation for all blocks (#2, #6, #7, #11, #13). Two subjects show no significant adaptation (#5, #8). The other 7 subjects show some adaptations depending on the considered parameters. This can be explained by the fact that subjects adopt different strategies. For example, subject #1 adapted his profile globally as shown by the significant variation of the 1^{st} moment. On the contrary, subject #12 principally adapted the beginning of the profile, as evidenced by the significant variation of the initial slope.

We performed a general statistical analysis (ANOVA) over the three blocks of the Adaptation phase for the 1^{st} moment and initial slope parameters, respectively left and

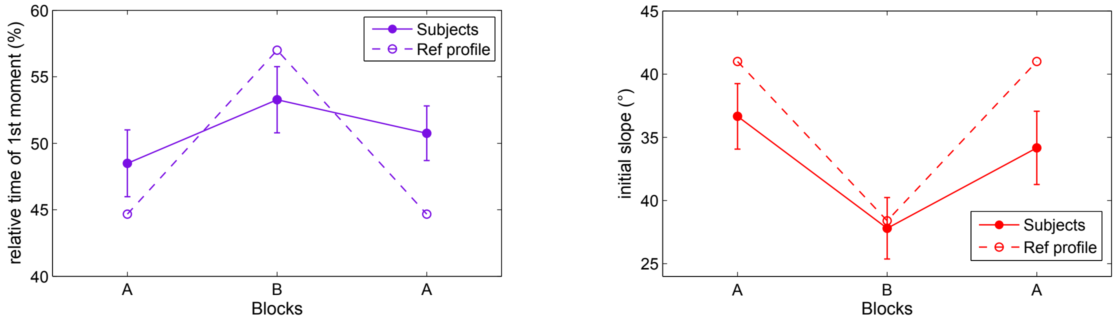


Figure 7.9 – Evolution of the relative time to 1^{st}moment (left) for the 14 last trials averaged for all subjects (plain lines) showing an underestimation of this parameter, and initial slope (right) showing the overshoot. Error bars indicate 95% confidence interval.

right on Figure 7.9. The analysis revealed a significant effect of the phase factor for both parameters: $F(2, 26) = 6.7, p < 0.005$ and $F(2, 26) = 11.5, p < 0.005$ respectively. Post-hoc tests indicated a significant change between transitions A-B and B-A for the 1^{st} moment and only for A-B transition for the initial slope. Therefore, these results show that subjects adapted their movement between the ends of each block, and this adaptation appeared more significant on 1^{st} moment. Interestingly, subjects tend to underestimate the adaptation of the 1^{st} moment position, indicating it might be difficult to modify that much this dynamic property of the movement. They also overshoot the initial slope towards small values, indicating they generally initiated their movement too slowly.

7.4.3 Qualitative Comments of the Subjects

The questionnaire filled by each subject offers additional information about the experiment. Concerning the Exploration phase, 8 subjects (out of 14) were positive that no change occurred in the system and 6 were unsure. Concerning the Adaptation phase, 8 subjects noticed that some changes occurred in the system, 5 were certain that no changes occurred, and 1 subject was convinced that the changes he perceived were solely due to his motion.

The subjects rated the difficulty of the task as 3.1 ± 0.9 and 3.1 ± 0.8 for the Exploration and Adaptation phases respectively (from 1-easy to 5-difficult). Subjects were also asked to evaluate the level of control they experienced over the system (from 1-no control at all, to 5-complete control). The results are close to the median: 2.3 ± 0.7 for the exploration phase and 2.8 ± 0.7 for the adaptation phase. Finally, they were asked questions concerning the system design. Subjects reported neither particular physical nor auditory fatigue (1.4 ± 0.6 and 1.2 ± 0.4 respectively, rated from 1 to 5). The perceived quality of the sounds produced was rated as 2.9 ± 0.9 over 5.

The performances of the subjects were not correlated with their sensation of control or success. Despite the fact that they all declared to be familiar with digital interfaces and practicing music (at a recreational level for most of them) we obtained quite heterogeneous results in terms of adaptation. It appears as though musical abilities or non-professional training is not a particular natural tendency to obtain better results in this particular adaptation task.

7.5 Discussion and Conclusion

We investigated the concept of *sound-oriented* task and questioned whether sound qualities could guide motion, and, in particular, its velocity profile. We proposed an experimental procedure to quantify how subjects adapt their gesture to produce a specific sound by avoiding either the presence of noise or of a loud higher harmonic.

Overall the results show that sensorimotor adaptations were found in both the Exploration and Adaptation experimental phases. In particular, 12 out of 14 subjects significantly adapted their movement to match the reference velocity profile during the Exploration phase. During the Adaptation phase, 12 out of 14 also showed some adaptation to the reference profiles, even if they were not informed of the sudden changes.

Nevertheless, important differences were noticed between subjects, which require further investigation. Several explanations can be put forward. First, participants, even musicians, appeared not to be used to manipulate sound with such a digital interface. The qualitative assessments of the subjects confirmed that the task was relatively difficult, which also indicates that the sensorimotor adaptation should be designed as more gradual. It is also noted that some subjects who obtained positive results did not notice the reference profiles variations. These observations are in favor of the presence of a strong implicit learning.

The type of extrinsic auditory feedback we developed in our experiment cannot be simply described using the well-known categories knowledge of result *KR* and knowledge of performance *KP* (Hartveld and Hegarty, 1996). Knowledge of result provides user with information on the success of the task, as a score for instance. Knowledge of performance provides user with information on the performance itself, such as information about kinematics or joint angles. In our case, the auditory feedback is used to adjust the angular velocity profile (faster or slower). In particular, it leads to corrections occurring in two steps, first during the motion to adapt it, and second, after the motion when planning for the next trial. The first role of the auditory feedback here is thus to provide information during the motion, which could be considered as KP. Nevertheless, the subjects also make use of the general auditory feedback during one trial in order to plan the next trial. The quantity of noise or harmonic they heard during a movement informs them on the success of their trial. Such a feedback could be considered to be similar to KR. This might explain why we did not observe a smooth improvement rate during the trials, but rather improvements based on trials and errors corrections.

From a sensorimotor loop perspective, the auditory feedback we propose is continuous. Moreover, the auditory feedback is designed as the task itself (as opposed to a motion-oriented task). This experiment served us as a first step to design more complete investigations of sensorimotor adaptation driven by *sound-oriented* tasks. In particular, it shows the limit of using standard feedback categories for continuous auditory feedback. Additional formal investigation and experimental works is thus necessary to establish a new framework to describe interactive auditory feedback and the regulation of the auditory-motor loop using these systems.

In conclusion, our results establish that learning movement kinematics is possible in an auditory task and allow us to support the notion of *sound-oriented* task for the study of sensorimotor learning. They open up towards new experiments which we are being pursued.

Acknowledgements

We acknowledge support by ANR French National Research Agency, under the ANR-Blanc program 2011 (Legos project ANR- 11-BS02-012) and additional support from Cap Digital. We thank all the participants of Legos project for very fruitful discussions.

Chapter 8

Discussion

8.1 Main results

The contribution of each study presented in our work is reported in this section.

Chapter 3

Pointing to virtual auditory targets allowed us to investigate the action-perception coupling in an auditory space. Although the final goal of pointing movements was to indicate a perceived direction, many results on perception and behavior were obtained, not only in the arm, but also in the head dynamics. The two conditions without movement sonification showed that temporal consistency between action and auditory feedback proved beneficial to guide pointing movement. Results showed that sound enables to use auditory representations of the environment and the body situation, and that these representations can be transformed into motor commands by the auditory-motor loop. Particularly, dynamic acoustic cues (here caused by head movements) are used by the auditory systems to process online motor control. This study highlights the role of this loop in spatial processing and the potential of auditory feedback for motor regulation. Besides, the design of the spatialized movement sonification turned out to be ineffective, certainly due to its similarity with the auditory targets.

Chapter 4

This study investigated whether a gesture-sound interaction can provide a learnable sensory substitution context. We developed the concept of Virtual Auditory Surfaces, where participants can interact through sound with a responsive region of space. The mapping ‘embodies’ the state and properties of the object and its interaction with hand motion. Experimental results showed that participants were able to determine, eyes closed, the curvature of a virtual plate using static as well as static and dynamic auditory feedback.

We also measured that the dynamical sonification, imaging the hand scratching the surface, increased the perceptual sensitivity and feelings experienced. This study illustrates that interactive sonification can allow for the perception of a virtual object and for building a representation of it. This representation, yet built without tangible interaction, proved to give rise to specific movements features (depending on the spatial context) and tactile-like sensations. Subjects reactions were equivalent to synaethesia feelings, as they ended up sensing the plate shape by how it sounded. The setup developed emphasizes the great potential of interplay between touch, proprioception and sound in interactive systems, and opens new perspectives for studying multimodal interaction with sound.

Chapter 5

We examined in this study the effects of continuous sonification in a standard visuo-manual tracking task. We showed that error-related, task-related as well as movement-related sonification with a simple design can improve tracking accuracy. No difference was found between the three feedback in terms of performance level increase. However, sonification of the user's motion proved to lessen the risks of developing a feedback dependency. In addition, it led to significant increase in performance and learning consolidation one day after training. One hypothesis can be that this auditory feedback is integrated as a supplementary proprioception of the arm dynamics, which, added to visual input, improves both the sensory feedback and prediction models in the sensorimotor loop. A major result of this study is the capacity of a continuous ecological sonification of tracking movement to heighten the energy of the motion. The fact that sonification can have a deep effect on movement features encourages the development of rehabilitation applications with this technique. Dynamics and temporal precision are essential parameters of designing sonification in the case of a visual task.

Chapter 6

This pilot study explored continuous sonification in a tracking task as well, but in the particular case of visuo-motor control. An experimental platform has been developed allowing for the sonification of ocular movements. A training protocol for learning free eye pursuit movements using a visual illusion has been proposed, along with visual tracking exercises. We first showed that sonification of smooth pursuits and saccades can increase tracking duration and gain in the case of the tracking task. Performance proved to vary importantly between individuals and stimuli patterns, suggesting modifications in the sonification process are needed. Secondly, when applied to free pursuit production with reverse-phi stimulus, sonification accelerated learning, and helped increasing the number of free eye pursuits. It also led to more energetic movements. Although all participants proved to react to sound, results vary largely between individuals as well. The study showed some cases of negative reaction to sonification, perceived as a counteractive feedback, especially when many saccades were produced. The presence of sonification sometimes led to feedback dependency. This study is an example where individual fine tuning of learning protocol and sonification is needed, because of the high difficulty of the task and the various stimuli parameters involved. Yet, it lays the cornerstone of using sonification to develop free smooth eye control and confirms the relevance of velocity-based and dynamical continuous movement sonification.

Chapter 7

In this chapter, we presented and developed the concept of sound-oriented task, inspired from musical control. It defines a motor task in a gestural sonic interaction context, where the goal of the task is expressed and presented as acoustic features. This design requires for the participants to memorize the auditory target; as a result, the mapping should be designed both without ambiguity and including enough dynamics and information content to enable movement correction. Results showed that participants were able to learn movement kinematics resulting in the targeted sound. Moreover, motor adaptation was observed when the mapping was changed with the same sound target (even to abnormal dynamics). However, the study showed that inter-individual variability was large in this framework. We confirmed that our concept represents a valuable framework for studying auditory-motor learning, enabling changes in the motor commands based on auditory-motor requirements. Proposing a task where movement is not seen as the purpose of the action proved to be engaging in spite of the task difficulty, and might allow for longer experiments with fewer weariness. This study raised again the diversity of behaviors observed when it comes to auditory-motor abilities.

8.2 General discussion

Our results show the general benefits of movement sonification on sensorimotor learning, whether it is supplementary to visual feedback or at the core of the gestural interaction. In particular, we demonstrate that gesture-sound interactive scenarios can be employed to create new sensorimotor learning situations. Reconsidering auditory feedback using continuous movement sonification allows for a “moving-while-listening” interactive context, favoring a specific sensorimotor coupling between sound and motion.

The choices of sounds and textures (mainly source-filter strategy), with dynamical amplitude and spectral content, can be debated, regarding aesthetics and pleasantness. In the end, this ecological choice, based on a physical ‘image’ of contact, friction and velocity, proved efficient on learning and influenced movement features: the sounds produced contained enough dynamical and spectral variations to provide significant and relevant information to the participants. Furthermore, no particular case of auditory fatigue was observed. The systems proposed in the sound-oriented task and the static feedback for the virtual surface experiment proved efficient as well. It seems that if the required information are present in the mapping and the resulting sonification, the choice of the sound texture might be less important. We would recommend to keep this 'direct' link between sound and kinematics in the interaction, in order to maximize energy in the movement for instance. This idea could be formalized as "ecological motion-sound mapping", *i.e.* continuous auditory feedback that relates to auditory-motor interaction we can be familiar with. Such a notion can be linked to the notion of affordance found in ecological psychology. It might be fruitful to examine gestural affordance favored by motion-sound mapping to help movement in a specific task or context.

We can nonetheless wonder about the need for a more metaphorical sound design. Some studies on movement have emphasized the 'understandability' of the movement sonification through accessible metaphors (Vogt et al., 2009). This supposes, conversely, that the users agree with the metaphor, or find it relevant for the task. Subjects' engagement could be amplified in such a case, especially in rehabilitation situations (see below). In the cases we studied, it seems that participants imagined (or built) their own metaphorical link between sound and motion, which was allowed by the relatively neutral sounds proposed.

Obviously, it is debatable whether ensuring an ecological (almost mechanical) link between sound and motion allows us to disregard any emotional influence on sensorimotor learning, as stated in the introduction of this work. Firstly, it is assuredly not possible to rule out completely emotional response, and secondly the emotional state can certainly impact positively or negatively sensorimotor learning. A part of the individual differences we observed might be due to a difference in emotional response (or condition). Many studies have also underlined the notion of reward for learning (Körding and Wolpert, 2006; Izawa and Shadmehr, 2011), sometimes based on emotional assumptions (Christensen, 2012).

As it has been discussed with the visuo-manual tracking study, imaging proprioception in the arm with sound might help for a robust learning and be unconsciously integrated into the sensorimotor loop. This would mean focusing the mapping on kinematic variables of the arm, like joint velocities, either in absolute values or relatively to other body segments (in the study, sonification was mapped on the pointer motion, that is the purpose of the movement). Another way to tighten the gesture-sound link could be to make use of the feeling of agency (Knoblich and Repp, 2009). On the one hand, agency is mainly studied with discrete and timed events (like tapping), and on the other hand it has been shown to reduced perceived exertion, but with musical stimuli (Fritz et al., 2013).

Improving learning by augmenting visual feedback with continuous sound seems possible, although it led to proportionally moderate performance improvements in our case. We believe that a difficult task would maximize the effects of the sonification, by offering larger room for improvement. It is the case in (Sigrist et al., 2014), where naive participants learned the complex task of rowing, or in our example of free eye pursuits generation, where sound effects can be quite large. As a side effect, in the case of a difficult or new task, inter-individual variability is also amplified. Secondly results showed that participants responded unequally to additional auditory feedback, highlighting a need to evaluate individual strategies of auditory versus visual feedback integration.

This large inter-individual variability has been observed generally in our studies. It appeared that some users are naturally responsive to continuous auditory-motor interactions, and some are not. We think that there might exist a predisposition to respond to these interactions, just like some people are not spontaneously good at playing football or clapping on a tempo. Few research addressed this question, although it has been suggested, in the (musical) case of learning to play a piano melody, that the speed of learning (hence the final performance results) is correlated with white matter fiber tracts, which connect brain areas associated with auditory-motor learning (Engel et al., 2014). The sound-oriented setup and particularly the virtual surface setup, where some participants did not succeed the task at all, might be suitable experimental contexts to evaluate this intrinsic capacity of people. This question should be further addressed if important applications are to be considered.

In chapter 7, we illustrated our concept of sound-oriented task and showed that an auditory-coded target paradigm could give rise to motor adaptation. All in all, this approach can be followed using many other setups and experiments: instead of describing the tasks with examples or semantics referring to movement features, an auditory reference is played and the auditory interaction is presented as the task itself. Memorizing visually a movement depicted as the target to reproduce (as it is often the case in sport, see below) can be difficult, because it can be composed of multiple sub-movements over many body segments and joints. On the contrary, we take advantage here of the capacity of our auditory system to memorize audio features in relatively short sequences (typically a few seconds). As each movement produces its typical sound, participants can learn implicitly this outcome, and be able to assess their performance and adjust their motor plans in terms of acoustic features after each trial. Continuous interaction with sound goes beyond the principles of auditory feedback, placing the participants actively in the action-perception loop. In addition, we argue that focusing the user's attention on auditory features can limit weariness, increase engagement and even make the task playful. These last points should be assessed in future research.

Results of chapters 3 and 4 showed that auditory representations of objects in space can arise from interactive sonification. The temptation is to use spatialized audio to augment spatial image of virtual objects (like in virtual reality setups). This would require advanced spatialization systems (which are not suitable for mobile applications) that need to be learned by the user, or new advances in near field binaural restitution (which is still investigated). Furthermore, the auditory perception under these conditions relies on subjective interpretation, which might worsen individual differences. A recent study showed indeed the difficulty of distance perception in peripersonal virtual environments (Parseihian et al., 2014). We believe that careful task definition, interactive design, and few spatialization concerns are enough to create an engaging gesture-sound interaction.

8.3 Limits

Beyond the encouraging results we obtained, this work also shows some limits in the experimental protocols or frameworks. First and foremost, as it has been emphasized by the diversity of behaviors observed, some of our studies could benefit from larger panels, especially the eye pursuit and sound-oriented task studies. In the latter, the relatively narrow panel yet allowed to observe a wide spectrum of behaviors, from very few motor adaptation to high performance levels. We can wonder whether increasing the number of participants in that case would have strengthened statistics or simply confirmed this diversity. The former was a pilot study, testing a particularly difficult motor acquisition, and which is currently being continued. In this study, performance evaluation relies on pursuit detection, based on kinematic and dynamic properties, which need to be precisely studied in the presence of reverse-phi stimulus. Furthermore, the amount of collected data require to formalize a robust analysis routine. Finally, fatigue, which seemed sometimes important, has not been deeply evaluated and should be taken into account for designing a new training protocol and to assess performance.

The variety of performance levels and behaviors observed during the experiments caused some difficulties in the analysis and interpretation of the results. This indicates that modeling this diversity might be necessary to achieve more robust statistical analysis, especially if larger panels are used. As evoked in the previous discussion, it seems that some results were blurred by the variety in the baseline 'auditory-motor capacities' of the participants, which has been highlighted through the different experiments.

An important aspect favoring the responsiveness could be the high level of training in music or sport activities practice. Although no significant effect were observed when trained participants were tested in our experiments, we hypothesize this training could lead to sizable effects, either in the auditory processing for the musicians, or in movement control and space representation for trained athletes. As it is illustrated in appendix A, trained athletes can exhibit very reproducible movement features, as motion is tuned around either a specific gesture or a specific outcome of the action. The question of assessing *a priori* these capacities is still under debate. The literature of pedagogical methods in music or sports can provide inspiring tracks for standardized auditory-motor tests, as well as rehabilitation research (Farrugia et al., 2012).

Retention tests after training periods with interactive sonification could have brought additional valuable information. However, a first and essential step of our work was to observe if and how continuous auditory feedback can be integrated in the sensorimotor loop and would affect direct performance. Longer studies including control groups and retention tests will be a direct continuation of this work.

8.4 Conclusive words

The experimental results we presented in this thesis can be summarized as a series of different principles to build distinct gesture-sound interactive systems, taking better advantage of auditory feedback. These points can also serve as starting points for novel experimental studies.

- Movement sonification can:
    - modify movement dynamics and improve task performance,
    - allow users to *listen* to specific movement features and integrate actively such additional sensory input for implicit movement adaptation and learning.
- Gesture-sound interaction and sonification can elicit the emergence of auditory-motor representations of actions and objects, and be used for sensory substitution.
- Sound-oriented tasks allow for transferring part of the user's attention from movement sensory information to the sound output. In this context, movement adaptation can be activated, driven only by sound.

Finally, there is a large variability among users skills in performing tasks with continuous auditory feedback. It is thus important that any sonic interactive system be adaptable to the user's skills. The development of a general framework to evaluate *a priori* such ability would be very helpful.

8.5 Perspectives

Summing up our different studies clearly shows the interest of continuous movement sonification for sensorimotor adaptation and learning, and that further studies are necessary to better understand underlying mechanisms. A substantial outcome of the work we presented is the identification of future perspectives, including important applications of movement sonification. Both research and out-of-the-lab contexts are targeted. Several experimental frameworks, like sensory substitution or oculomotor control, can benefit from interactive movement sonification. The originality and the appealing results of the sound-oriented task study support a new promising framework for motor control and multisensory integration studies in interactive sonic contexts (auditory-visual and auditory-haptic). As it as been previously evoked, sonification of oculomotor behavior for learning free smooth pursuits is another experimental perspective. The potential application to training protocols for learning eye writing for ALS patients is an important motivation.

Movement sonification could also be further applied to proprioception and tactile perception studies. The idea of augmenting proprioception with sound represents a new field of investigation for the use of sonification. As we hypothesized in the tracking study, continuous sonification can be used to give an auditory image of proprioception, either while moving body parts or when manipulating an object. The spontaneous comments of participants in the virtual surface experiment, reporting tactile sensations, or even heat, when interacting with the virtual surface, should be further investigated. These cross-modal effects of movement sonification should first be confirmed and exploited to enrich gestural sonic interactions. An example of application field can be virtual reality, which is growing fast and is always in demand for multisensory integration. De-afferent patients could also be tested in these systems to investigate the capacity of sensory substitution of sound.

Throughout this work, we developed movement sonification systems with straightforward and rather simple audio techniques, for experimental control reasons, but also with a view of applications for different populations. The encouraging results we obtained can be pushed forward taking advantage of portable and cost-effective motion capture systems and audio systems that can be embedded in mobile platforms (including high end audio techniques like for instance concatenative synthesis of textures). Furthermore, we hope our findings could benefit the design of control interfaces using these technologies, integrating sensorimotor aspects for better 'learnability'.

Sport stands as a particular connected field for studying auditory-motor coupling. As we highlighted that sonifying an error, relatively to a reference gesture, might be difficult, we argue that the use of sonification for sport should turn towards the exploration of "moving-while-listening". The same way athletes more or less consciously learned to use acoustic outcomes of their actions when practicing, we believe movement sonification can be used for implicit learning of fine dynamic movement features. A systematic approach in table tennis has been initiated (cf annexe A) and will be continued, both in professional and recreational practice. The challenge is not to define the gesture-sound mapping for a reference gesture, but to express relevant movement features with sound; that is, going from the 'error sonification' strategy to a sound-oriented task.

This idea of moving-while-listening intends to highlight the special relationship between action and auditory perception, and encourages ecological movement sonification for a new vision of auditory 'feedback'. One of the important perspectives of this work is the development of sensorimotor studies where sound is a central element, not just a feedback. Among the potential applications already cited, physical rehabilitation is one of the most promising.

Physical rehabilitation and movement sonification

The motivations and background literature for using sonification in physical rehabilitation have been detailed in section 2.3.4. As a reminder, 80% of stroke survivors are affected by hemiparesis[1]. Many results are available in the literature, although with various success, methods and pathologies addressed (mainly stroke survivors and Parkinson's patients). We emphasize here that continuous sonification, providing more than knowledge of result on movements, has been rarely used in rehabilitation of the arm after stroke. We recently started investigating this usage, with the will to redesign the approach of feedback in this discipline (Roby-Brami et al., 2014). Specifically, the coordination between shoulder, elbow and wrist is generally affected after surviving a stroke, and compensation movements with the whole trunk leaning forward are often observed in reaching movements (Roby-Brami et al., 2003). We propose to investigate sonification methods for the coordination of arm reaching movements. During a preliminary work, we developed two approaches inspired from the sound-oriented task to apply sonification techniques to arm rehabilitation exercises. The first one (depicted in an extended abstract available in appendix B) consisted in mapping the time profile of shoulder/elbow coordination in a reaching movement to the synthesis of textures using metaphorical content. It constitutes a first step towards a sound-oriented task paradigm for rehabilitation exercises of arm coordination.

A second approach aims at observing motor adaptation driven by audio features introducing, this time, a motor perturbation in the movement using a controllable-break orthosis. Similarly to the first approach, an audio target is defined in a kinematic parameters space. Preliminary tests are currently being carried out to evaluate this setup.

For patients in rehabilitation process, simple everyday motor tasks become difficult and they present large room for improvement. Additionally, movements affected and recovery progression can be very different from one patient to another. It seems that the guidelines we draw about task difficulty and individual differences could also apply in rehabilitation. Stimulating energy in motion and augmenting proprioception with sound can also be a promising path for patients.

These two experimental strategies open the way towards a *sonic interaction approach of physical rehabilitation*. Alongside with technical and sound designs, important aspects about, motivation, engagement and playfulness should be tested as well, where interactive sound can be a valuable asset. Emphasizing the motivational aspect, the elderly population could benefit from physical exercises proposed in the "moving-while listening" context. The development of interactive sonic objects (with affordances driven by sound) can address children with autistic syndromes, proposing multisensory experiences during therapy.

Future projects, inviting physical doctors, physiotherapists, orthosis and robotics specialists will address this application, which we believe to be a major outcome of our work.

---

[1] According to the National Stroke Association `http://www.stroke.org/`

Appendices and supplementary articles

Appendix A

Low-cost motion sensing of table tennis players for real time feedback

E. O. Boyer[1,3], F. Bevilacqua[1], F. Phal[2], S. Hanneton[2,3]
[1] IRCAM, STMS-CNRS-UPMC, Paris, France
[2] UFR STAPS, Université Paris Descartes, Paris, France
[3] LPP, Paris Descartes CNRS UMR 8242, Paris, France

Refereed article published in *International Journal of Table Tennis Sciences*, 8, 2013.

Abstract: We present a motion capture device to measure in real-time table tennis strokes. A six degree-of-freedom sensing device, inserted into the racket handle, measures 3D acceleration and 3-axis angular velocity values at a high sampling rate. Data are wirelessly transmitted to a computer in real-time. This flexible system allows for recording and analyzing kinematics information on the motion of the racket, along with synchronized video and sound recordings. Recorded gesture data are analyzed using several algorithms we developed to segment and extract movement features, and to build a reference motion database.

Keywords: Sonification, Movement analysis, Biomechanics, Sensors, Motor learning, Auditory feedback, Gesture, Sound.

A.1 Introduction

There is a growing demand for visualizing and analyzing data in sports sciences whether it is for professional athletes to enhance their performance, educational applications or entertainment. We present here a motion capture device we applied to table tennis that was originally developed for gestural control of sound and music. Most studies dealing with table tennis techniques and players movement analysis used video analysis (Rodrigues et al., 2002), high-speed cameras for slow motion (Shum and Komura, 2005) or motion capture systems (Rusdorf and Brunnett, 2005). Kinematics data such as velocities and accelerations were therefore estimated based on position and displacement with differential calculus (Bootsma et al., 2010). We propose here to use inertial sensors to measure and analyze players' movements, namely three-dimensional accelerometers and gyroscopes. Such sensors have been integrated in tennis rackets and golf club (Perkins, 2007; Telford, 2006) for sensing acceleration rates and swing speed.

A.2 Motion sensing system

The motion sensing system we developed is based on wireless inertial measurement units (IMU) embedded in the rackets. IMU are nowadays available at low cost and can measure angular velocity (gyroscope) and acceleration (accelerometer) at sufficiently high sampling rate. Compared to motion capture systems based on camera, IMU are largely superior in terms of cost and ease of use for our application where fast and relatively small wrist motions must be analyzed in real-time.

Accelerometers measure a value that contains two components: a 'dynamic' acceleration corresponding to 2$^{\text{nd}}$ derivative of the position relative to a given axis, and a 'static' acceleration which depends on the orientation of the sensors to the gravity. The proper acceleration may thus differ from the coordinate acceleration, which is calculated by derivation on spatial coordinates and displacement. A gyroscope is a device that measures an angular velocity around a rotation axis. Three-axis gyros deliver angular velocities around x, y and z axis.

IMU necessitates the need to modify the racket to embed the sensors in the handle. This is relatively easy to achieve for experiments without modifying significantly the shape and weights of the rackets. In particular, we used a small version of wireless motion sensing modules called mini-MO that was designed for digital music instruments (Fléty and Maestracci, 2011; Rasamimanana et al., 2011). Such modules were used to equip all sort of musical instruments such a violin bow (Kimura et al., 2012) or everyday objects like sport balls (Rasamimanana et al., 2012).

Technically, the mini-MO contains a 3D accelerometer (Analog Device ADXL345) and a 3-axis gyroscope (InvenSense ITG-3200). These sensors are integrated in a custom electronic card (4×3 cm) that also contains a module (Jennic OEM JN-5139) that process and transmit the data wirelessly using the IEEE protocol (2.4 GHz band). The unit is powered using a rechargeable battery. Figure A.1 shows a picture of our prototype.

Y axis is parallel to the racket handle directed toward the top of the racket, Z is normal to the racket surface directed out from the red covering and X is perpendicular to the racket handle oriented so that (X,Y,Z) is a direct cartesian coordinate system.

Data are transmitted to the host computer running a recording software programmed

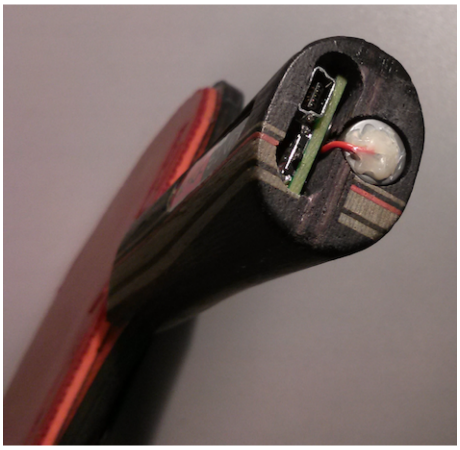


Figure A.1 – View of a racket prototype with a mini-MO and its battery inserted in the handle.

with Max/MSP (`www.cycling74.com`) and using UDP and OSC (`www.opensoundcontrol.org`) protocols. Data were recorded at 10 ms sample rate along with video and sound of the performance. Video was recorded with an IEEE-1394 camera (Pointgrey Research, FL2-03S2C) at 30 frames per second. Ambient sound of ball bounces on the table and strokes on the racket surface was recorded using a AKG CK91 capsule on a AKG SE300B condenser microphone in an Edirol UA-25 audio interface running at 44.1 kHz.

A.3 Experiment

A.3.1 Subjects

Two table tennis players, graduating for Sport Sciences and Technology degree at Université Paris Descartes (Paris, France) participated in the first experiment using the system described in the previous section. They were both 20 years old and both right-handed. A unique session was carried out under normal training session conditions and with the help of their coach.

A.3.2 Experimental procedure

The aim of this preliminary study was to build a first database of gesture and to analyze the movement features that can be obtained with our system. The players were asked to perform repetitive strokes with their coach sending them a ball without spin during practice (no rally). The first player was asked to perform six specific strokes: forehand and backhand drive, forehand and backhand topspin and forehand and backhand push. The second player was asked to perform a forehand and a backhand service. Players were instructed to play until they feel satisfied with at least 10 strokes. We recorded all the gestures continuously. Between 22 and 50 strokes were recorded for each type and 16 forehand and backhand services each.

A.4 Results

A dedicated interface using the software Max/MSP was developed for visualizing and playing back all data streams synchronously (i.e. six-degree motion data from the sensors, sound and video tracks). A screenshot of the interface is presented in figure A.2. The playback of the gestures allows for simulating real-time data coming from the sensors and building up real-time processes like controlling sounds for example. Moreover, by moving the cursor or jumping to a precise time in the recording, one can browse over the different strokes recorded and play back specific excerpts. Data, video and sound tracks are synchronized together at a position that can be controlled by the cursor, which enable for detailed observations of the player's gesture.

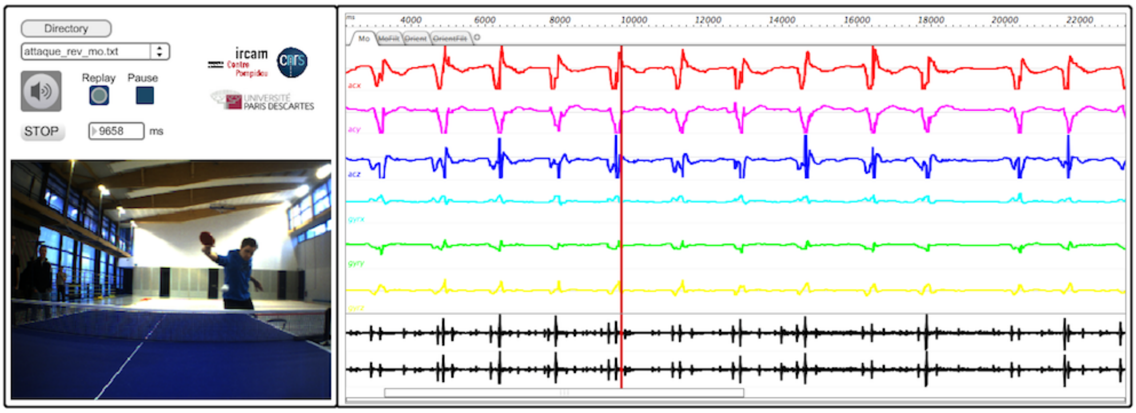


Figure A.2 – Screenshot of the interface developed for visualizing the measurements and aligned media tracks. The top six curves are accelerations and angular velocities for X, Y and Z axis respectively. The waveform of the audio recording is visible at the bottom. The vertical cursor allows scrolling forward and backward while all tracks are played synchronously. Control panel is visible at top left and video window at bottom left.

It is possible to detect to moment the ball hits the racket surface (Sorensen et al., 2001) directly from the raw acceleration signal. An example of raw acceleration data is presented in figure A.3 where a sharp peak can be observed, which corresponds to the shock produced by the ball hitting the racket surface. Note that the shock appears, as expected, synchronously on the 3 axis. The amplitude on these peaks appears to be typically two to four times the maximum amplitude of the acceleration signals corresponding to the racket motion. The peak of the ball hit can be easily removed using simple low-pass signal filtering since it has a high frequency content compared to the gesture.

One purpose of constituting a database of players' movements is to find procedures to evaluate the movement repeatability. All raw data were processed with a Gaussian low-pass filter to remove stroke peaks. The complete recordings, containing repetitive strokes, were segmented by finding initiation and termination of the movements in the synchronized video track. Figure A.4 shows the superposition of the 37 forehand drive strokes (in gray lines) for the 6 data channels (top graphs from the left are acceleration on X, Y and Z axis, and bottom graphs are angular velocities around X, Y and Z axis). For each stroke

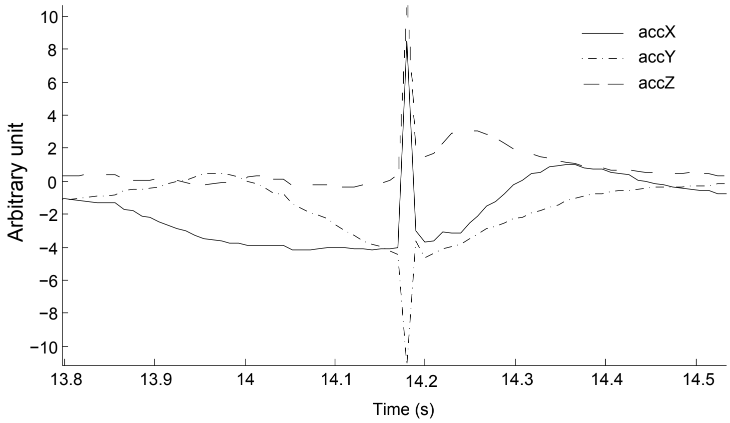


Figure A.3 – Raw acceleration time profile on X, Y and Z axis of a forehand drive showing the effect on acceleration of the ball hitting the racket surface.

the black circle indicates the moment of the contact with the ball. The average across all the movements is plotted in plain black line. The two dash lines represent the mean curve plus or minus the standard deviation. Amplitudes are normalized.

Overall, we found that there is a high repeatability of the player motion, which is illustrated in Figure A.4 by the small standard deviation, especially on X and Y accelerations. The Z angular velocity exhibits smaller and more constant standard deviation over time compared to the X and Y angular velocities. Generally, X and Y gyros data revealed smaller magnitude and slightly higher standard deviation across repetition. This is actually not surprising considering that for this particular stroke, angular velocities of the racket around these axes are not prevailing.

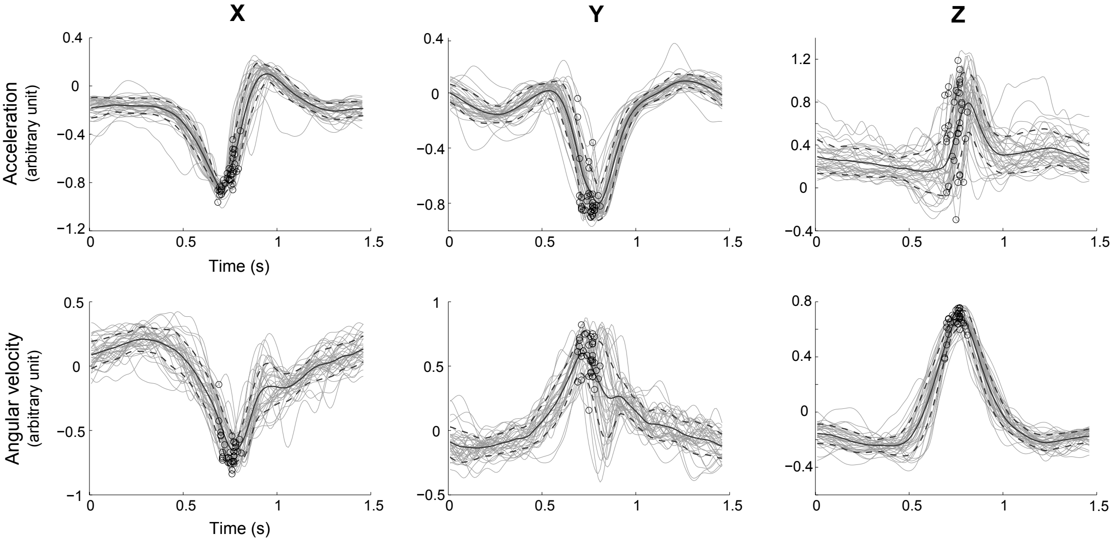


Figure A.4 – 3D acceleration (top row) and angular velocities (bottom row) of the 37 consecutive forehand drive strokes, black circles indicating ball hits. Black bold line is the average; dash lines are the average plus and minus standard deviation.

The evolution over time of the standard deviation shows a drop near the ball contact with the racket. This reveals that the player movements are 'tuned' around the time of contact with the ball. The player seems to adapt his motion to strike the ball each time in the most consistent manner (depending on his position and the coming ball spin and trajectory).

Our results are concordant with Marinovic (Marinovic et al., 2004) who studied with video cameras the velocity of forehand drive strokes under several condition changes (trajectory, velocity and spin of the ball). They concluded that players adjusted their movement toward striking the ball with a maximum velocity. In our study the players actually stroke the ball around maxima and minima of acceleration and angular velocity of the racket. In the case of a forehand drive, the ball hits appear on average 40 ms after the minimum on X axis, and 50 ms before minimum and maximum on Y and Z axis, respectively.

The evolution of the standard deviation observed in the forehand and backhand services revealed smaller values than for the other strokes (typically twice as less). This could be explained by the greater control of the whole motion, since the ball is initially thrown by the player himself, giving a better control on the whole stroke. Less motor adaptation is then required during service, producing reduced variation in the data.

A.5 Conclusion and perspectives

This study demonstrates that our motion sensing system, using inertial sensors, is sufficiently precise to study typical tennis table strokes. Our low-cost system can retrieve required information about movement kinematics and detect the ball strokes on the racket. The interface developed stands for a complete multimedia visualization tool available right after a recording without processing needed. Replay also allows for recreating real-time conditions for prototyping further implementations. A preliminary database of gestures has been built up and will be extended to other types of strokes (Tepper, 2013) and in different conditions.

This first database aims at creating a framework for developing sonification of player's movements. Our purpose is to give the player an augmented feedback of his movement using the auditory modality. Our hypothesis is that motor control can be supported by an additional sensory feedback (Todorov et al., 1997) and that sound can be an effective and pleasant way to provide information fast enough to embody the player's movement (Karageorghis and Terry, 1997; Effenberg, 2007; Boyer et al., 2013). Table tennis players already constantly integrate information about sensory inputs, mainly vision and proprioception, with extremely fast processing and triggering of motor commands. The auditory modality remains largely available without interfering with other modalities and can be processed rapidly. Therefore, sound can be an effective way to provide extra-information in real-time to the player during practice. Combining movement analysis with sonification techniques can provide players with direct information for self-assessment of the quality of strokes or for the intention of the opponent. This allows for new perspectives of the representation of the players' motion.

A.6 Acknowledgments

This work was supported by the ANR (French National Research Agency) Legos project (11 BS02 012). We thank Emmanuel Fléty and Alain Terrier for their precious help in building the prototypes.

Appendix B

Sonification of the coordination of arm movements

F. Bevilacqua[1], A. Van Zandt-Escobar[1,2], N. Schnell[1], E. O. Boyer[1,3], N. Rasamimanana[4], J. Françoise[1], S. Hanneton[3], A. Roby-Brami[2]
[1] IRCAM, STMS-CNRS-UPMC, Paris, France
[2] ISIR, UPMC, UMR CNRS 7222, Paris, France
[3] LPP, Paris Descartes CNRS UMR 8242, Paris, France
[4] Phonotonic, France

Refereed extended abstract of *Multisensory Motor Behavior: Impact of Sound*, 2013.

Abstract: We investigate sonification of arm movements for the rehabilitation of stroke patients. This involves providing patients with auditory feedback relative to the coordination between shoulder and elbow motion, and relative to movement smoothness. To this purpose, we are exploring different types of sonification and musical metaphors, including source -filter, concatenative-granular and physical modeling sound synthesis methods.

B.1 Introduction

We are currently experimenting various approaches to providing auditory feedback related to the quality of the arm movement. Sonification of human movement is a growing research topic, with early results showing showing promise in a wide range of applications, such as performing arts, rehabilitation and sports training (Hermann et al., 2011; Rasamimanana et al., 2011). We are concerned with the specific case of patients recovering from strokes. These patients generally suffer from important motor impairments that alter the normal coordination of their arm movement (Roby-Brami et al., 2003), and diminish movement smoothness. A continuous auditory feedback mechanism, linked to the Knowledge of Performance (KP), might be beneficial in the rehabilitation process (Cirstea and Levin, 2007). In particular, we propose here to associate the movements to specific continuous sound qualities that should be "targeted" by the patients. We report here a cost-effective system that implements different sonification strategies.

B.2 Motion sensing and analysis

The setup is illustrated in figure B.1 left. The arm is equipped with two wireless sensing modules containing Inertial Motion Units with 6 DOF (accelerometer and gyroscope) (Rasamimanana et al., 2011). The data is streamed at a frame rate of 200Hz and is analyzed by custom software built with Max6 (Cycling'74). Different types of real-time data analysis are available, including:

- Computation of the $\theta(t)$ and $\beta(t)$ angles, and their relative angular velocities (figure B.1).
- Various smoothness parameters based on peaks detection, zero-crossing, or Fourier analysis of angular velocities and acceleration data.

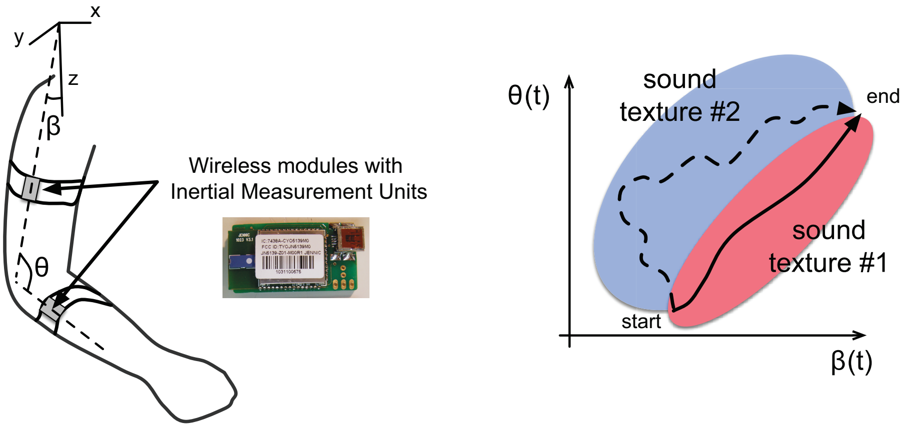


Figure B.1 – Left: sensing system for the arm motion. Right: Trajectories $\theta(t)$ and $\beta(t)$ mapped to different continuously parameterized sound textures.

B.3 Sonification strategies

Multiple sonification strategies are implemented, utilizing different sound synthesis techniques. These different cases are studied experimentally in order to establish pros and cons of each technique. In particular, the following synthesis methods are used:

- Source-filter synthesis.
- Concatenative and granular synthesis[1].
- Physical modeling sound synthesis.

---

[1] MuBu software, see `http://forumnet.ircam.fr`

The motion parameters are mapped to synthesis parameters using either ‘direct mapping’ or intermediate models (Hermann et al., 2011; Rasamimanana et al., 2011). In some cases, the motion/sound relationships can be perceived as ‘metaphorical’. Figure B.1 right illustrates one of the cases we implemented. The time profiles $\theta(t)$ and $\beta(t)$ can be represented as 2D trajectories that traverse different regions, which are mapped to different sound textures. These textures are controlled with granular or concatenative sound synthesis, using recorded sounds that are automatically segmented, analyzed and labeled.

Acknowledgment ANR (French National Research Agency) Legos project (11 BS02 012) and Labex SMART (supported by French state funds managed by the ANR within the ‘Investissements d’Avenir’ program under reference ANR-11-IDEX-0004-02).

Appendix C

Technical specifications of the eyetracker

SR Research
EyeLink

SR Research
Complete Eye Tracking Solutions

EyeLink 1000 Technical specifications

	EyeLInk 1000					
	Tower	Primate	*Desktop*	Arm	Long Range	Remote Option
Monocular Sampling Rate	250, 500, 1000, 2000 Hz					250, 500 Hz
Binocular Sampling Rate		250, 500, 1000 Hz			250, 500, 1000 Hz	
Eye Tracking Principle	Pupil with CR	Pupil Only Pupil with CR	Pupil with Corneal Reflection(CR)			
Average Accuracy	0.25° to 0.5° typical					0.5° typical
Saccade Event Resolution	0.05° microsaccades					0.25°
Spatial Resolution (RMS)	0.01° @ 1000 Hz 0.02° @ 2000 Hz					0.05°
End to End Sample Delay	M < 1.8 msec, SD < 0.6 msec @ 1000 Hz M < 1.4 msec, SD < 0.4 msec @ 2000 Hz					M < 3.0 msec, SD=1.11 msec
Blink Recovery Time	1.0 msec @ 1000 Hz 0.5 msec @ 2000 Hz					2 msec @ 500 Hz
Pupil Detection Models	Centroid or Ellipse Fitting					Ellipse Fitting
Gaze Tracking Range	60° horizontally, 40° vertically		32° horizontally 25 ° vertically			
Allowable Head Movement	25x25x10 mm (horizontal x vertical x depth)					22x18x20 cm (horizontal x vertical x depth)
Optimal Camera-Eye Distance	38 cm	30-45 cm	40 - 70 cm		60 - 150 cm	40 - 70 cm
Glasses Compatability	Good		Excellent			Good
Infrared Wavelength	910 nm or 940 nm		890 nm or 940 nm			

Bibliography

Alonso-Arevalo, M., Shelley, S., Hermes, D., Hollowood, J., Pettitt, M., Sharples, S., and Kohlrausch, A. (2012). Curve shape and curvature perception through interactive sonification. *ACM Transactions on Applied Perception*, 9(4):1–19. 82

Altenmüller, E. O. (2003). Focal dystonia: advances in brain imaging and understanding of fine motor control in musicians. *Hand Clinics*, 19(3):523–538. 35

Anderson, M. L. (2003). Embodied Cognition: A field guide. *Artificial Intelligence*, 149(1):91–130. 20

Auvray, M., Hanneton, S., and O'Regan, K. J. (2007). Learning to perceive with a visuo – auditory substitution system: Localisation and object recognition with 'The vOICe'. *Perception*, 36(3):416–430. 70

Avanzini, F., De Götzen, A., Spagnol, S., and Rodà, A. (2009). Integrating auditory feedback in motor rehabilitation systems. In *Proceedings of International Conference on Multimodal Interfaces for Skills Transfer (SKILLS09)*. 41, 48, 89, 137

Avanzini, F., Rocchesso, D., and Serafin, S. (2004). Friction sounds for sensory substitution. In *Proceedings the 10th International Conference on Auditory Display - ICAD 2004*, pages 1–8. 71, 136

Avanzini, F., Spagnol, S., Rodà, A., and de Götzen, A. (2011). Designing interactive sound for motor rehabilitation tasks. In *Sonic Interaction Design*. MIT Press. 42, 43, 48

Aytekin, M., Moss, C. F., and Simon, J. Z. (2008). A sensorimotor approach to sound localization. *Neural computation*, 20(3):603–35. 20, 65

Bach-y Rita, P. (1970). Neurophysiological basis of a tactile vision-subsitution system. *IEEE Transactions on Man-Machine Sytems*, 11(1):108–110. 14, 68

Bach-y Rita, P. and Baillet, R. (1987). Recovery from stroke. In *Stroke rehabilitation: The recovery of motor control*, page 98. Year Book Medical. 46

Bach-y Rita, P., Collins, C. C., Saunders, F. A., White, B., and Scadden, L. (1969). Vision Substitution by Tactile Image Projection. *Nature*, 221(5184):963–964. 68

Bach-y Rita, P. and W. Kercel, S. (2003). Sensory substitution and the human–machine interface. *Trends in Cognitive Sciences*, 7(12):541–546. 69, 84

Baier, G., Hermann, T., and Stephani, U. (2007). Event-based sonification of EEG rhythms in real time. *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, 118(6):1377–86. 22, 43

Baram, Y. and Miller, A. (2007). Auditory feedback control for improvement of gait in patients with Multiple Sclerosis. *Journal of the neurological sciences*, 254(1-2):90–4. 4, 46

Basta, D., Singbartl, F., Todt, I., Clarke, A., and Ernst, A. (2008). Vestibular rehabilitation by auditory feedback in otolith disorders. *Gait & posture*, 28(3):397–404. 46

Bastian, A. J. (2010). Understanding sensorimotor adaptation and learning for rehabilitation. *Current opinion in neurology*, 21(6):628–633. 16

Batavia, M., Gianutsos, J. G., and Kambouris, M. (1997). An Augmented auditory feedback device. *Archives of physical medicine and rehabilitation*, 78(December). 46

Batavia, M., Gianutsos, J. G., Vaccaro, A., and Gold, J. T. (2001). A do-it-yourself membrane-activated auditory feedback device for weight bearing and gait training: a case report. *Archives of physical medicine and rehabilitation*, 82(4):541–5. 47

Baudry, L., Leroy, D., Thouvarecq, R., and Choller, D. (2006). Auditory concurrent feedback benefits on the circle performed in gymnastics. *Journal of Sports Science*, 24(2):149–156. 39, 40

Baumann, O. and Greenlee, M. W. (2009). Effects of attention to auditory motion on cortical activations during smooth pursuit eye tracking. *PloS one*, 4(9):e7110. 108

Berryhill, M. E., Chiu, T., and Hughes, H. C. (2006). Smooth pursuit of nonvisual motion. *Journal of neurophysiology*, 96(1):461–5. 106, 107

Berthoz, A. (2003). *La décision.* Odile Jacob, Paris. 20

Bevilacqua, F., Fels, S., Jensenius, A. R., Lyons, M. J., Schnell, N., and Tanaka, A. (2013a). SIG NIME. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems on - CHI EA '13*, page 2529, New York, New York, USA. ACM Press. 35, 133

Bevilacqua, F., Fléty, E., Schnell, N., Caramiaux, B., Rasamimanana, N., Françoise, J., Bloit, J., and Boyer, E. O. (2013b). De-MO : Designing Action-Sound Relationships with the MO Interfaces. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, pages 2907–2910. 6

Bevilacqua, F., Müller, R., and Schnell, N. (2005). MnM : a Max / MSP mapping toolbox. In *New Interfaces for Musical Expression.* 112

Bevilacqua, F., Schnell, N., and Fdili Alaoui, S. (2011). Gesture capture: Paradigms in interactive music/dance systems. *Emerging Bodies: The Performance of Worldmaking in Dance and Choreography*, 183:183–193. 38

Boian, R., Deutsch, J., Lee, C., Burdea, G., and Lewis, J. (2003). Haptic effects for virtual reality-based post-stroke rehabilitation. In *11th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems, 2003. HAPTICS 2003.*, pages 247–253. 49

Bologna, G., Deville, B., and Pun, T. (2009). On the use of the auditory pathway to represent image scenes in real-time. *Neurocomputing*, 72(4-6):839–849. 70

Bootsma, R. J., Fernandez, L., Morice, A. H. P., and Montagne, G. (2010). Top-level players' visual control of interceptive actions: Bootsma and van Wieringen (1990) 20 years later. *Journal of Experimental Psychology: Human Perception and Performance*, 36(4):1056–63. 162

Bouchara, T., Katz, B. F. G., Guastavino, C., and Jacquemin, C. (2010). Audio-visual Renderings For Multimedia Navigation. In *Proceedings of the 16th International Conference on Auditory Display ICAD-2010*, Washingont, D.C., USA. 31

Bovermann, T., Groten, J., Campo, A. D., and Eckel, G. (2007). Juggling Sounds. In *Proceedings of the 2nd International Workshop on Interactive Sonification*, pages 1–6, York, UK. 40

Boyer, E. O., Babayan, B. M., Bevilacqua, F., Noisternig, M., Warusfel, O., Roby-Brami, A., Hanneton, S., and Viaud-Delmon, I. (2013). From ear to hand: the role of the auditory-motor loop in pointing to an auditory source. *Frontiers in computational neuroscience*, 7(April):26. 70, 135, 166

Boyer, E. O., Pyanet, Q., Hanneton, S., and Bevilacqua, F. (2014). Learning Movement Kinematics with a Targeted Sound. In *Sound, Music & Motion, Lecture Notes in Computer Science Volume 8905*, pages 218–233. Springer International Publishing. 88, 100

Brainard, M. S. and Doupe, A. J. (2000). Auditory feedback in learning and maintenance of vocal behaviour. *Nature Reviews Neuroscience*, 1:31–40. 35

Brashers-Krug, T., Shadmehr, R., and Todorov, E. (1995). Catastrophic Interference in Human Motor Learning. In *Advances in Neural Information Processing Systems*, volume 7. MIT press Cambridge, Massachusetts, Cambridge, MA, USA. 17

Bravi, R., Del Tongo, C., Cohen, E. J., Dalle Mura, G., Tognetti, A., and Minciacchi, D. (2014). Modulation of isochronous movements in a flexible environment: links between motion and auditory experience. *Experimental Brain Research*, 232(6):1663–75. 34

Bridgeman, B. (1995). A review of the role of efference copy in sensory and oculomotor control systems. *Annals of Biomedical Engineering*, 23(4):409–422. 109

Brock, H., Schmitz, G., Baumann, J., and Effenberg, A. O. (2012). If motion sounds: Movement sonification based on inertial sensor data. *Procedia Engineering*, 34:556–561. 23

Brown, R. M. and Palmer, C. (2013). Auditory and motor imagery modulate learning in music performance. *Frontiers in human neuroscience*, 7(July):320. 33

Brungart, D. S., Simpson, B. D., Mckinley, R. L., Kordik, A. J., Dallman, R. C., and Ovenshire, D. A. (2004). The interaction between head-tracker latency, source duration, and response time in the localization of virtual sound sources. In *Proceedings of the 10th International Conference on Auditory Display ICAD 04*, pages 1–7. 53

Buchanan, J. J. and Wang, C. (2012). Overcoming the guidance effect in motor skill learning: feedback all the time can be beneficial. *Experimental Brain Research*, 219(2):305–20. 29, 89, 98, 130

Bulkin, D. A. and Groh, J. M. (2006). Seeing sounds: visual and auditory interactions in the brain. *Current opinion in neurobiology*, 16(4):415–9. 4

Cameirao, M. S., Bermuidez, S., Zimmerli, L., Oller, E. D., and Verschure, P. F. M. J. (2007). The Rehabilitation Gaming System : a Virtual Reality Based System for the Evaluation and Rehabilitation of Motor Deficits. In IEEE, editor, *Proceedings of Virtual Rehabilitation*, pages 29–33, Venice, Italy. 48, 49, 84

Caramiaux, B. (2011). *Études sur la Relation Geste–Son en Performance Musicale.* PhD thesis, Université Pierre et Marie Curie, Paris, France. 19, 136

Caramiaux, B., Bevilacqua, F., Bianco, T., Schnell, N., Houix, O., and Susini, P. (2014). The Role of Sound Source Perception in Gestural Sound Description. *ACM Transactions on Applied Perception*, 11(1):1–19. 23, 136

Castiello, U., Giordano, B. L., Begliomini, C., Ansuini, C., and Grassi, M. (2010). When ears drive hands: the influence of contact sound on reaching to grasp. *PloS one*, 5(8):e12240. 25

Chen, J. L., Penhune, V. B., and Zatorre, R. J. (2009). The role of auditory and premotor cortex in sensorimotor transformations. *Annals of the New York Academy of Sciences*, 1169:15–34. 32, 52

Chiari, L., Dozza, M., Cappello, A., Horak, F. B., Macellari, V., and Giansanti, D. (2005). Audio-Biofeedback for Balance Improvement: An Accelerometry-Based System. *IEEE Transactions on Biomedical Engineering*, 52(12):2108–2111. 47

Christensen, E. (2012). *Music Listening, Music Therapy, Phenomenology and Neuroscience.* PhD thesis, Aalborg University, Denmark. 152

Cirstea, C. M. and Levin, M. F. (2007). Improvement of arm movement patterns and endpoint control depends on type of feedback during practice in stroke survivors. *Neurorehabilitation and neural repair*, 21(5):398–411. 170

Cirstea, C. M., Ptito, A., and Levin, M. F. (2006). Feedback and cognition in arm motor skill reacquisition after stroke. *Stroke; a journal of cerebral circulation*, 37(5):1237–42. 42

Collicutt, M., Casciato, C., and Wanderley, M. M. (2009). From real to virtual: A comparison of input devices for percussion tasks. In *Proccedings of the International Conference on New Interfaces for Musical Expression (NIME) 2009*, pages 4–6. 38

Collins, C. (1970). Tactile Television-Mechanical and Electrical Image Projection. *IEEE Transactions on Man-Machine Systems*, 11(1):65–71. 68

Collins, N. (2011). LL : Listening and Learning in an Interactive Improvisation System. Technical report, University of Sussex. 38

Colombo, R., Pisano, F., Micera, S., Mazzone, A., Delconte, C., Carrozza, M. C., Dario, P., and Minuco, G. (2005). Robotic techniques for upper limb evaluation and rehabilitation of stroke patients. *IEEE transactions on neural systems and rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society*, 13(3):311–24. 48

Conde, V., Altenmüller, E. O., Villringer, A., and Ragert, P. (2012). Task-irrelevant auditory feedback facilitates motor performance in musicians. *Frontiers in Psychology*, 3(May):146. 33

Correa, A. G. D., de Assis, G. A., do Nascimento, M., Ficheman, I., and de Deus Lopes, R. (2007). GenVirtual: An Augmented Reality Musical Game for Cognitive and Motor Rehabilitation. *Virtual Rehabilitation*, pages 1–6. 49, 84

Correia, V., Araujo, D., Craig, C., and Passos, P. (2011). Prospective information for pass decisional behavior in rugby union. *Human Movement Science*, 30:984–997. 39

Craik, K. (1947). Theory of the human operator in control systems. I. The operator as an engineering system. *British Journal of Psychology*, 38(2):56–61. 88

Dalla Bella, S., Giguère, J.-F., and Peretz, I. (2009). Singing in congenital amusia. *The Journal of the Acoustical Society of America*, 126(1):414–24. 35

Danna, J., Fontaine, M., Paz-Villagrán, V., Gondre, C., Thoret, E., Aramaki, M., Kronland-Martinet, R., Ystad, S. l., and Velay, J.-L. (2014). The effect of real-time auditory feedback on learning new characters. *Human Movement Science*, In press. 26

Danna, J., Paz-Villagrán, V., and Velay, J. L. (2013a). Signal-to-Noise velocity peaks difference: A new method for evaluating the handwriting movement fluency in children with dysgraphia. *Research in Developmental Disabilities*, 34:4375–4384. 26

Danna, J., Velay, J.-L., Paz-Villagrán, V., Capel, A., Petroz, C., Gondre, C., Thoret, E., Aramaki, M., Ystad, S., and Kronland-Martinet, R. (2013b). Handwriting movement sonification for the rehabilitation of dysgraphia. In *10th International Symposium on Computer Music Multidisciplinary Research (CMMR) - Sound, Music & Motion*, pages 200–208, Marseille, France. 26

Darainy, M., Vahdat, S., and Ostry, D. J. (2013). Perceptual learning in sensorimotor adaptation. *Journal of neurophysiology*, 110(9):2152–62. 18, 100, 131

Davidson, P. R. and Wolpert, D. M. (2004). Scaling down motor memories: de-adaptation after motor learning. *Neuroscience letters*, 370(2-3):102–7. 16, 17, 131

de Götzen, A., Rocchesso, D., and Serafin, S. (2006). Does Fitts' law sound good? In *3rd International Conference on Enactive Interfaces (Enactive /06)*, volume 2, pages 123–124. 24

De Lucia, M., Camen, C., Clarke, S., and Murray, M. M. (2009). The role of actions in auditory object discrimination. *NeuroImage*, 48(2):475–85. 4

Dozza, M., Horak, F. B., and Chiari, L. (2007). Auditory biofeedback substitutes for loss of sensory information in maintaining stance. *Experimental Brain Research*, 178(1):37–48. 4, 43, 45

Dubus, G. (2012). Evaluation of four models for the sonification of elite rowing. *Journal on Multimodal User Interfaces*, 5:143–156. 27

Dubus, G. (2013). *Interactive sonification of motion.* PhD thesis, KTH University of Stockholm. 1, 22, 30

Dubus, G. and Bresin, R. (2011). Sonification of Physical Quantities Throughout History : A Meta-Study Of Previous Mapping Strategies. In *The 17th International Conference on Auditory Display.* 30

Dubus, G. and Bresin, R. (2013). A systematic review of mapping strategies for the sonification of physical quantities. *PloS one*, 8(12):e82491. 30

Dubus, G. and Bresin, R. (2014). Exploration and evaluation of a system for interactive sonification of elite rowing. *Sports Engineering*, September. 27

Dufour, A., Touzalin, P., and Candas, V. (2007). Rightward shift of the auditory subjective straight ahead in right- and left-handed subjects. *Neuropsychologia*, 45(2):447–53. 70

Easton, R. D., Greene, a. J., DiZio, P., and Lackner, J. R. (1998). Auditory cues for orientation and postural control in sighted and congenitally blind people. *Experimental Brain Research*, 118(4):541–50. 45

Eckhouse, R. H., Morash, R. P., and Maulucci, R. (1990). Sensory Feedback and the Impaired Motor System. *Journal of Medical Systems*, 14(3):93–105. 44

Effenberg, A. O. (2004). Using Sonification to Enhance Perception and Reproduction Accuracy of Human Movement Patterns. In *Proceedings of the International Workshop on Interactive Sonification*, Bielefeld. 24, 88, 136

Effenberg, A. O. (2007). Movement Sonification : Motion perception, behavioral effects and functional data. In *Internation Workshop on Interactive Sonification*, pages 1–4. 166

Effenberg, A. O., Fehse, U., and Weber, A. (2011). Movement Sonification: Audiovisual benefits on motor learning. *BIO Web of Conferences*, 1:00022. 40, 136

Effenberg, A. O., Weber, A., and Melzer, J. (2005). MotionLab Sonify : A Framework for the Sonification of Human Motion Data. In *9th International Conference on Information Visualisation*, number 1. 24

Engardt, M. (1994). Long term effects of auditory feedback training on relearned symmetrical body weight distribution in stroke patients. *Scandinavian Journal of Rehabilitation Medecine*, 26:65–69. 17, 42, 46

Engel, A., Hijmans, B. S., Cerliani, L., Bangert, M., Nanetti, L., Keller, P. E., and Keysers, C. (2014). Inter-individual differences in audio-motor learning of piano melodies and white matter fiber tract architecture. *Human brain mapping*, 35:2483–97. 153

Engel, A. K., Maye, A., Kurthen, M., and König, P. (2013). Where's the action? The pragmatic turn in cognitive science. *Trends in cognitive sciences*, 17(5):202–9. 20

Farrugia, N., Benoit, C.-E., Harding, E., Kotz, S. A., and Dalla Bella, S. (2012). BAASTA : Battery for the Assessment of Auditory Sensorimotor and Timing Abilities. In *Proceedings of the 12th International lConference on Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music*, Thessaloniki, Greece. 35, 154

Ferguson, S. (2006). Learning Musical Instrument Skills Through Interactive Sonification. In *Proccedings of the International Conference on New Interfaces for Musical Expression (NIME) 2006*, pages 384–389. 39

Fine, I. and Jacobs, R. a. (2002). Comparing perceptual learning tasks: a review. *Journal of vision*, 2(2):190–203. 131

Fitts, P. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 47(6). 24

Flash, T. and Hogan, N. (1985). The Coordination of Arm Movements: An Experimentally Confirmed Mathematical Model. *The Journal of neuroscience*, 5(7):1688–1703. 15

Fléty, E. and Maestracci, C. (2011). Latency improvement in sensor wireless transmission using IEEE 802.15. 4. In *International Conference on New Interfaces for Musical Expression (NIME)*, pages 409–412. 162

Forma, V., Hoellinger, T., Auvray, M., Roby-Brami, A., and Hanneton, S. (2011). Ears on the hand: reaching 3D audio targets. *BIO Web of Conferences*, 1:00026. 16, 27, 52, 136

Fox, J. and Carlile, J. (2005). SoniMime : Movement Sonification for Real-Time Timbre Shaping. In *Proccedings of the International Conference on New Interfaces for Musical Expression (NIME) 2005*, pages 242–243, Vancouver, BC, Canada. 38

Françoise, J., Schnell, N., and Bevilacqua, F. (2013a). A multimodal probabilistic model for gesture–based control of sound synthesis. *Proceedings of the 21st ACM international conference on Multimedia - MM '13*, pages 705–708. 22

Françoise, J., Schnell, N., and Bevilacqua, F. (2013b). Gesture-based control of physical modeling sound synthesis. In *Proceedings of the 21st ACM international conference on Multimedia - MM '13*, pages 447–448, New York, New York, USA. ACM Press. 22

Franinović, K. and Serafin, S. (2012a). *Sonic Interaction Design - Part 1.* MIT Press. 22, 31

Franinović, K. and Serafin, S. (2012b). *Sonic Interaction Design - Part 2.* MIT Press. 22

Franklin, D. W. and Wolpert, D. M. (2011). Computational mechanisms of sensorimotor control. *Neuron*, 72(3):425–42. 15, 16, 17

Fritz, T. H., Hardikar, S., Demoucron, M., Niessen, M., Demey, M., Giot, O., Li, Y., Haynes, J.-D., Villringer, A., and Leman, M. (2013). Musical agency reduces perceived exertion during strenuous physical performance. *Proceedings of the National Academy of Sciences of the United States of America*, pages 1–6. 40, 42, 152

Gauthier, G. and Hofferer, J. (1976). Eye tracking of self-moved targets in the absence of vision. *Experimental Brain Research*, 26(2):121–139. 107

Gaver, W. W. (1993a). How Do We Hear in the World?: Explorations in Ecological Acoustics. *Ecological Psychology*, 5(4):285–313. 75

Gaver, W. W. (1993b). What in the World Do We Heatr?: An Ecological Approach to Auditory Event Perception. *Ecological Psychology*, 5(1):1–29. 22, 75, 100

Geiger, C., Reckter, H., Paschke, D., Schulz, F., and Poepel, C. (2008). Towards participatory design and evaluation of theremin-based musical interfaces. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME 2008)*, pages 303–306, Genova, Italy. 38

Gelineck, S. and Serafin, S. (2009). A Quantitative Evaluation of the Differences between Knobs and Sliders. In *Proccedings of the International Conference on New Interfaces for Musical Expression (NIME) 2009.* 135

Geronazzo, M., Avanzini, F., and Grassi, M. (2013). Influence of auditory pitch on haptic estimation of spatial height. In *Proceedings of the 10th International Symposium on Computer Music Interdisciplinary Research, Marseille, France, October 15-18*, pages 758–765. 31

Ghez, C., Rikakis, T., Dubois, R. L., and Cook, P. R. (2000). An Auditory Display System for Aiding Interjoint Coordination. In *Proceedings of International Conference on Auditory Display.* 45

Gibet, S. (2009). Sensorimotor control of sound-producing gestures. In *Musical Gestures: Sound, Movement and Meaning.* Routledge. 36

Gibson, J. J. (1986). *The ecological approach to visual perception.* Psychology Press. 19

Godøy, R. I., Haga, E., and Jensenius, A. R. (2006a). Exploring Music-Related Gestures by Sound-Tracing . A Preliminary Study . In *Proceedings of the COST287-ConGAS 2nd International Symposium on Gesture Interfaces for Multimedia Systems (GIMS2006).* 136

Godøy, R. I., Haga, E., and Jensenius, A. R. (2006b). Playing "Air Instruments": Mimicry of Sound-producing Gestures by Novices and Experts. In Gibet, S., Courty, N., and Kamp, J.-F., editors, *Gesture in Human-Computer Interaction and Simulation.*, volume 3881 of *Lecture Notes in Computer Science*, pages 256–267. Springer Berlin Heidelberg, Berlin, Heidelberg. 136

Godøy, R. I. and Leman, M. (2010). *Musical gestures: Sound, movement, and meaning.* Routledge. 36

Gonzalez, J., Soma, H., Sekine, M., and Yu, W. (2012). Psycho-physiological assessment of a prosthetic hand sensory feedback system base on an auditory display: a preliminary study. *Journal of neuroengineering and rehabilitation*, 9(1):33. 45

Goossens, H. H. L. M. and Van Opstal, A. J. (1999). Influence of Head Position on the Spatial Representation of Acoustic Targets. *Journal of neurophysiology*, 81:2720–2736. 64

Grosshauser, T. and Hermann, T. (2009). The sonified music stand - an interactive sonification system for musicians. In *Proceedings of the 6th Sound and Music Computing Conference*, pages 233–238, Porto, Portugal. 39

Grosshauser, T. and Hermann, T. (2010). Multimodal Closed-loop Human Machine Interaction. In *Proceedings of the 3rd International workshop on Interactive Sonification.* 31

Guitton, D. (1992). Control of eye-head coordination during orienting gaze shifts. *Trends in neurosciences*, 15(5):174–179. 63

Hanneton, S., Auvray, M., and Durette, B. (2010). The Vibe : a versatile vision-to-audition sensory substitution device. *Applied Bionics and Biomecahnics*, 7(4):269–276. 70

Hanneton, S., Berthoz, A., Droulez, J., and Slotine, J. (1997). Does the brain use sliding variables for the control of movements? *Biological Cybernetics*, 77(6):381–393. 88, 99

Harris, C. M. and Wolpert, D. M. (1998). Signal-dependent noise dertermines motor planning. *Nature*, 394(20). 15

Hartveld, A. and Hegarty, J. (1996). Augmented Feedback and Physiotherapy Practice. *Physiotherapy*, 82(8):480–490. 12, 13, 42, 88, 146

Hasan, Z. (1986). Optimized movement trajectories and joint stiffness in unperturbed, inertially loaded movements. *Biological Cybernetics*, 53(6):373–382. 15

Hebb, D. O. (1949). *The organization of behavior: A neuropsychological approach.* John Wiley & Sons. 14

Hermann, T., Hunt, A., and Neuhoff, J. G. (2011). *The Sonification Handbook.* Logos Verlag, Berlin. 22, 31, 169, 171

Hickok, G., Buchsbaum, B., Humphries, C., and Muftuler, T. (2003). Auditory-motor interaction revealed by fMRI: speech, music, and working memory in area Spt. *Journal of cognitive neuroscience*, 15(5):673–82. 35, 52

Hoffmann, C. P., Moens, B., Dalla Bella, S., and Bardy, B. G. (2013). Does running in synchrony with sound improve endurance performance and save energy? In *Proceedings of the 10th International Symposium on Computer Music Interdisciplinary Research, Marseille, France, October 15-18*, pages 158–162. 40

Hoffmann, C. P., Torregrosa, G., and Bardy, B. G. (2012). Sound Stabilizes Locomotor-Respiratory Coupling and Reduces Energy Cost. *PLoS ONE*, 7. 40

Holden, M. K. (2005). Virtual Environments for Motor Rehabilitation: Review. *Cyberpsychology & behavior*, 8(3):187–211. 49, 84

Holland, S., Bouwer, A. J., Dalgelish, M., and Hurtig, T. M. (2010). Feeling the beat where it counts. *Proceedings of the fourth international conference on Tangible, embedded, and embodied interaction - TEI '10*, page 21. 38

Hornof, A. (2014). The Prospects For Eye-Controlled Musical Performance. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME) 2014*, pages 461–466. 108

Houde, J. F. and Jordan, M. I. (1998). Sensorimotor adaptation in speech production. *Science*, 279(5354):1213–6. 35

Hsu, W. and Sosnick, M. (2009). Evaluating Interactive Music Systems : An HCI Approach. In *Proccedings of the International Conference on New Interfaces for Musical Expression (NIME) 2009*, pages 25–28. 38

Huang, H., Ingalls, T., Olson, L., Ganley, K., Rikakis, T., and He, J. (2005). Interactive multimodal biofeedback for task-oriented neural rehabilitation. *27th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, 3:2547–50. 137

Hunt, A. and Kirk, R. (2000). Mapping Strategies for Musical Performance. In Wanderley, M. M. and Battier, M., editors, *Trends in Gestural Control of Music*, pages 231–258. Ircam - Centre Pompidou. 38, 135

Hunt, A., Wanderley, M. M., and Kirk, R. (2000). Towards a Model for Instrumental Mapping in Expert Musical Interaction. In *Proceedings of the 2000 International Computer Music Conference*, pages 209–212. 38

Hunt, A., Wanderley, M. M., and Paradis, M. (2003). The importance of parameter mapping in electronic instrument design. *Journal of New Music Research*, 32(4):429–440. 38

Izawa, J. and Shadmehr, R. (2011). Learning from sensory and reward prediction errors during motor adaptation. *PLoS Computational Biology*, 7(3):e1002012. 152

Jenkins, W., Merzenich, M., Ochs, M., Allard, T., and Guic-Robles, E. (1990). Functional Reorganization of Primary Somatosensory Cortex in Adult Owl Monkeys After Behaviorally Controlled Tactile Stimulation. *Journal of neurophysiology*, 63(1):82–104. 69

Jesteadt, W. (1980). An adaptive procedure for subjective judgments. *Perception & psychophysics*, 28(1):85–8. 74

Johnson, M. J., Van der Loos, H. F. M., Burgar, C. G., Shor, P., and Leifer, L. J. (2003). Design and evaluation of Driver's SEAT: A car steering simulation environment for upper limb stroke therapy. *Robotica*, 21(01). 49

Kagerer, F. A. and Contreras-Vidal, J. L. (2009). Adaptation of sound localization induced by rotated visual feedback in reaching movements. *Experimental Brain Research*, 193(2):315–21. 16, 17, 88

Karageorghis, C. I. and Terry, P. C. (1997). The psychophysical effects of music in sport and exercice : a review. *Journal of Sport Behavior*, 20(1):54. 40, 42, 137, 166

Katz, B., Warusfel, O., Bourdot, P., and Vezien, J.-M. (2007). CoRSAIRe – Combination of Sensori-motor Rendering for the Immersive Analysis of Results. In *2nd Internation Workshop on Interactive Sonification*, number 1, pages 1–4, York, UK. 31

Kerlirzin, Y., Dietrich, G., and Vieilledent, S. (2009). *Le Contrôle moteur.* Presses Universitaires de France. 12, 15, 18

Kerzel, D., Born, S., and Souto, D. (2010). Inhibition of steady-state smooth pursuit and catch-up saccades by abrupt visual and auditory onsets. *Journal of neurophysiology*, 104(5):2573–85. 108

Kiefer, C., Collins, N., and Fitzpatrick, G. (2008). HCI methodology for evaluating musical controllers: A case study. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME 2008)*. 38

Kim, J., Hashida, T., Ohtani, T., and Naemura, T. (2012). Effects of auditory feedback for augmenting the act of writing. *Proceedings of the 3rd Augmented Human International Conference on - AH '12*, pages 1–4. 25

Kimura, M., Rasamimanana, N., Bevilacqua, F., Schnell, N., Zamborlin, B., and Fléty, E. (2012). Extracting human expression for interactive composition with the augmented violin. In *International Conference on New Interfaces for Musical Expression (NIME)*, pages 99–102, Ann Arbor. 162

Knoblich, G. and Flach, R. (2001). Predicting the effects of actions: interactions of perception and action. *Psychological science*, 12(6):467–72. 20

Knoblich, G. and Repp, B. H. (2009). Inferring agency from sound. *Cognition*, 111(2):248–62. 40, 152

Kohler, E., Keysers, C., Umiltà, M. A., Fogassi, L., Gallese, V., and Rizzolatti, G. (2002). Hearing sounds, understanding actions: action representation in mirror neurons. *Science (New York, N.Y.)*, 297(5582):846–8. 4

Körding, K. P. and Wolpert, D. M. (2006). Bayesian decision theory in sensorimotor control. *Trends in cognitive sciences*, 10(7):319–26. 152

Krakauer, J. W. (2006). Motor learning: its relevance to stroke recovery and neurorehabilitation. *Current opinion in neurology*, 19(1):84–90. 69

Kramer, G., Walker, B. N., Bonebright, T., Cook, P. R., Flowers, J., and Miner, N. (1999). Sonification Report: Status of the Field and Research Agenda. Technical report, University of Nebraska - Lincoln. 22

Krauzlis, R. J. (2005). The control of voluntary eye movements: new perspectives. *The Neuroscientist : a review journal bringing neurobiology, neurology and psychiatry*, 11(2):124–37. 104, 106

Lang, A. (2011). *Rôle des régularités perceptives dans l'apprentissage implicite d'un mouvement répété*. PhD thesis, UTC Compiègne. 16

Larkin, O., Koerselman, T., Ong, B., and Ng, K. (2008). Sonification of bowing features for string instrument training. In *ICAD 2008*, pages 2–5. 39

Lee, D. N., Craig, C. M., and Grealy, M. A. (1999). Sensory and intrinsic coordination of movement. *Proceedings. Biological sciences / The Royal Society*, 266:2029–2035. 19

Lee, D. N., Georgopoulos, A. P., Clark, M. J., Craig, C. M., and Port, N. (2001). Guiding contact by coupling the taus of gaps. *Experimental Brain Research*, 139:151–159. 20

Lee, J.-Y. and Schweighofer, N. (2009). Dual adaptation supports a parallel architecture of motor memory. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 29(33):10396–404. 16

Lehrer, N., Attygalle, S., Wolf, S. L., and Rikakis, T. (2011a). Exploring the bases for a mixed reality stroke rehabilitation system, Part I: a unified approach for representing action, quantitative evaluation, and interactive feedback. *Journal of neuroengineering and rehabilitation*, 8(1):51. 49, 84

Lehrer, N., Chen, Y., Duff, M., L Wolf, S., and Rikakis, T. (2011b). Exploring the bases for a mixed reality stroke rehabilitation system, Part II: design of interactive feedback for upper limb rehabilitation. *Journal of neuroengineering and rehabilitation*, 8(1):54. 49, 84

Lemaitre, G., Houix, O., Visell, Y., Franinović, K., Misdariis, N., and Susini, P. (2009). Toward the design and evaluation of continuous sound in tangible interfaces: The Spinotron. *International Journal of Human-Computer Studies*, 67(11):976–993. 31, 32

Lemaitre, G. and Susini, P. (2009). The Flops Glass : A Device To Study Emotional Reactions Arising From Sonic Interactions. In *SMC 2009 - 6th Sound and Music Conference*, pages 1–6, Porto, Portugal. 31

Leman, M. (2008). *Embodied music cognition and mediation technology.* Mit Press. 19, 33

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical Society of America*, 49(2):Suppl 2:467+. 74

Lisberger, S. G., Morris, E. J., and Tychsen, L. (1987). Visual motion processing and sensory-motor integration for smooth pursuit eye movements. *Annual review of neuroscience*, 10:97–129. 106

Lorenceau, J. (2012). Cursive writing with smooth pursuit eye movements. *Current biology : CB*, 22(16):1506–9. 104, 106, 107

Lorenz, R. H., Berndt, A., Groh, R., and Dresden, R. (2013). Designing Auditory Pointers. In *Proceedings of the 8th Audio Mostly Conference*, page 6. ACM. 31

Loureiro, R. U. I., Amirabdollahian, F., Topping, M., and Driessen, B. (2003). Upper Limb Robot Mediated Stroke Therapy — GENTLE / s Approach. *Autonomous Robots*, 15(1):35–51. 41, 49

Madelain, L. and Krauzlis, R. J. (2003). Effects of learning on smooth pursuit during transient disappearance of a visual target. *Journal of neurophysiology*, 90(2):972–82. 108

Maes, P.-J., Leman, M., Palmer, C., and Wanderley, M. M. (2014). Action-based effects on music perception. *Frontiers in psychology*, 4(January):1008. 33

Maier, J. X. and Groh, J. M. (2009). Multisensory guidance of orienting behavior. *Hearing research*, 258(1-2):106–12. 4, 65

Makous, J. C. and Middlebrooks, J. C. (1990). Two-dimensional sound localization by human listeners. *The Journal of the Acoustical Society of America*, 87(5):2188–200. 64, 70

Malloch, J., Birnbaum, D., Sinyor, E., and Wanderley, M. M. (2006). Towards a new conceptual framework for digital musical instruments. In *Proceedings of the 9th International Conference on Digital Audio Effects*, pages 49–52. 37

Marinovic, W., Iizuka, C. A., and Freudenheim, A. M. (2004). Control of striking velocity by table tennis players. *Perceptual and motor skills*, 99(3 Pt 1):1027–34. 166

Masiero, S. and Carraro, E. (2013). Upper limb movements and cerebral plasticity in post-stroke rehabilitation. *Aging Clinical and Experimental Research*, 20(2):103–108. 14

Mathias, B., Palmer, C., Perrin, F., and Tillmann, B. (2014). Sensorimotor Learning Enhances Expectations During Auditory Perception. *Cerebral cortex (New York, N.Y. : 1991)*. 33

Maulucci, R. a. and Eckhouse, R. H. (2001). Retraining reaching in chronic stroke with real-time auditory feedback. *NeuroRehabilitation*, 16(3):171–82. 17, 41, 44

Mcintosh, G. C., Brown, S. H., Rice, R. R., and Thaut, M. H. (1997). Rhytmic auditory-motor facilitation of gait patterns in patients with Parkinson's disease. *Journal of Neurology, Neurosurgery and Psychiatry*, 62:22–26. 47

McIntyre, J., Stratta, F., and Lacquaniti, F. (1998). Short-term memory for reaching to visual targets: psychophysical evidence for body-centered reference frames. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 18(20):8423–35. 62

McRuer, D. (1980). Human dynamics in man-machine systems. *Automatica*, 16:237–253. 88

Merer, A., Aramaki, M., Ystad, S. l., and Kronland-Martinet, R. (2013). Perceptual characterization of motion evoked by sounds for synthesis control purposes. *ACM Transactions on Applied Perception*, 10(1):1–24. 23

Merians, A. S., Poizner, H., Boian, R., Burdea, G., and Adamovich, S. (2006). Sensorimotor training in a virtual reality environment: does it improve functional recovery poststroke? *Neurorehabilitation and neural repair*, 20(2):252–67. 84

Merleau-Ponty, M. (1945). *Phénoménologie de la perception.* Gallimard, Paris. 19

Merrill, D. J. and Paradiso, J. A. (2005). Personalization, expressivity, and learnability of an implicit mapping strategy for physical interfaces. In *CHI 2005 Conference on Human Factors in Computing Systems*, page 2152. Citeseer. 38

Meyer, C., Lasker, A., and Robinson, D. (1985). The upper limit of human smooth pursuit velocity. *Vision research*, 25(4):6–8. 112

Miall, R. C., Weir, D. J., and Stein, J. F. (1993). Intermittency in human manual tracking tasks. *Journal of motor behavior*, 25(1):53–63. 88

Miall, R. C. and Wolpert, D. M. (1996). Forward Models for physiological motor control. *Neural networks*, 9(8):1265–1279. 20, 100

Miranda, E. R. and Wanderley, M. M. (2006). *New digital musical instruments: control and interaction beyond the keyboard.* AR Editions. 35, 36

Mitrovic, D., Klanke, S., Osu, R., Kawato, M., and Vijayakumar, S. (2010). A Computational model of limb impedance control based on principles of internal model uncertainty. *PLoS One*, 5:e13601. 138

Münte, T. F., Altenmüller, E., and Jäncke, L. (2002). The musician's brain as a model of neuroplasticity. *Nature reviews. Neuroscience*, 3(6):473–478. 32

Nef, T., Mihelj, M., Kiefer, G., Perndl, C., Muller, R., and Riener, R. (2007). ARMin - Exoskeleton for Arm Therapy in Stroke Patients. In *2007 IEEE 10th International Conference on Rehabilitation Robotics*, pages 68–74. IEEE. 49

Norman, J. F., Kappers, A. M. L., Cheeseman, J. R., Ronning, C., Thomason, K. E., Baxter, M. W., Calloway, A. B., and Lamirande, D. N. (2013). Aging and curvature discrimination from static and dynamic touch. *PloS one*, 8(7):e68577. 82

Novembre, G. and Keller, P. E. (2014). A conceptual review on action-perception coupling in the musicians' brain: what is it good for? *Frontiers in human neuroscience*, 8(August):603. 33

Nymoen, K., Glette, K., Torresen, J., and Jensenius, A. R. (2010). Searching for Cross-Individual Relationships between Sound and Movement Features using an SVM Classifier. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME 2010)*, pages 259–262, Sydney, Australia. 136

Nymoen, K., Stravinsky, P. I., and Torresen, J. (2011). Analyzing Sound Tracings - A Multimodal Approach to Music Information Retrieval. In *Proceedings of the 1st international ACM workshop on Music information retrieval with user-centered and multimodal strategies*, pages 4–9. 136

O'Modhrain, S. (2011). A Framework for the Evaluation of Digital Musical Instruments. *Computer Music Journal*, 35(1):28–42. 38

O'Modhrain, S. and Essl, G. (2004). PebbleBox and CrumbleBag: tactile interfaces for granular synthesis. In *Proceedings of the 2004 Conference on New Interfaces for Musical Expression*, pages 74–79. 32

O'Regan, K. J. (2011). *Why red doesn't sound like a bell.* Oxford University Press. 69

O'Regan, K. J. and Noë, A. (2002). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24(05):939–973. 20

Oscari, F., Secoli, R., Avanzini, F., Rosati, G., and Reinkensmeyer, D. J. (2012). Substituting auditory for visual feedback to adapt to altered dynamic and kinematic environments during reaching. *Experimental Brain Research*, 221(1):33–41. 27, 52

Ostry, D. J., Darainy, M., Mattar, A. a. G., Wong, J. D., and Gribble, P. L. (2010). Somatosensory plasticity and motor learning. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 30(15):5384–93. 17

Paillard, J. (1985). Les niveaux sensori-moteur et cognitif du contrôle de l'action. In *Recherches en Activités Physiques et Sportives*, pages 174–163. Centre de Recherche de l'UEREPS, Université Aix-Marseille II. 12

Paine, G. (2009). Towards unified design guidelines for new interfaces for musical expression. *Organised Sound*, 14(2):142–155. 36

Paine, G. (2010). Towards a taxonomy of realtime interfaces for electronic music performance. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, pages 436–439. 36

Palmer, C., Koopmans, E., Loehr, J. D., and Carter, C. (2009). Movement-related Feedback and temporal accuracy in clarinet performance. *Music Perception*, 26(5):439–450. 33, 85

Paraskevopoulos, I. T., Tsekleves, E., Craig, C., Whyatt, C., and Cosmas, J. (2014). Design guidelines for developing customised serious games for Parkinson's Disease rehabilitation using bespoke game sensors. *Entertainment Computing*, 5:413–424. 47

Parseihian, G. (2012). *Sonification binaurale pour l'aide à la navigation.* PhD thesis, Université Pierre et Marie Curie, Paris. 70

Parseihian, G., Jouffrais, C., and Katz, B. F. G. (2014). Reaching nearby sources: comparison between real and virtual sound and visual targets. *Frontiers in neuroscience*, 8(September):269. 153

Pascual-Leone, A., Amedi, A., Fregni, F., and Merabet, L. B. (2005). The plastic human brain cortex. *Annual review of neuroscience*, 28:377–401. 14

Pauletto, S. and Hunt, A. (2009). Interactive sonification of complex data. *International Journal of Human-Computer Studies*, 67(11):923–933. 22, 43

Perkins, N. C. (2007). Electronic measurement of the motion of a moving body of sports equipment. US Patent 7,234,351. 162

Perrott, D. R., Ambarsoom, H., and Tucker, J. (1987). Changes in head position as a measure of auditory localization performance: auditory psychomotor coordination under monaural and binaural listening conditions. *The Journal of the Acoustical Society of America*, 82(5):1637–45. 64

Pfordresher, P. Q. and Bella, S. D. (2011). Delayed auditory feedback and movement. *Journal of Experimental Psychology: Human Perception and Performance*, 37(2):566. 34

Pinek, B. and Brouchon, M. (1992). Head turning versus manual pointing to auditory targets in normal subjects and in subjects with right parietal damage. *Brain and cognition*, 18(1):1–11. 64

Poepel, C. (2005). On interface expressivity: a player-based study. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME 2005)*, pages 228–231. 38

Pouget, A., Ducom, J. C., Torri, J., and Bavelier, D. (2002). Multisensory spatial representations in eye-centered coordinates for reaching. *Cognition*, 83(1):B1–B11. 65

Prablanc, C., Pélisson, D., and Goodale, A. (1986). Visual control of reaching movements without vision of the limb. *Experimental Brain Research*, 62:293–302. 62

Proulx, M. J. (2010). Synthetic synaesthesia and sensory substitution. *Consciousness and cognition*, 19(1):501–3. 14, 69

Rasamimanana, N. and Bevilacqua, F. (2008). Effort-based analysis of bowing movements: evidence of anticipation effects. *Journal of New Music Research*, 37(4):339–351. 36

Rasamimanana, N., Bevilacqua, F., Bloit, J., Schnell, N., Fléty, E., Cera, A., and Frechin, P. (2012). The urban musical game: using sport balls as musical interfaces. In *CHI'12 Extended Abstracts on Human Factors in Computing Systems*, pages 1027–1030. 162

Rasamimanana, N., Guédy, F., Schnell, N., Lambert, J.-P., and Bevilacqua, F. (2008). Three pedagogical scenarios using the sound and gesture lab. In *Proceedings of the 4th i-Maestro Workshop on Technology Enhanced Music Education*, Genova, Italy. 39

Rasamimanana, N., Schnell, N., Fléty, E., and Zamborlin, B. (2011). Modular Musical Objects Towards Embodied Control Of Digital Music Real Time Musical Interactions. In *Procedings of TEI'11*, pages 1–4. 162, 169, 170, 171

Rasmussen, J. (1983). Skills , Rules , and Knowledge ; Signals , Signs , and Symbols , and Other Distinctions in Human Performance Models. *IEEE Transactions on Systems, Man and Cybernetics*, 13(3):257–266. 37

Rath, M. and Schleicher, R. (2008). On the Relevance of Auditory Feedback for Quality of Control in a Balancing Task. *Acta Acustica united with Acustica*, 94:12–20. 26, 32, 88, 136

Rizzolatti, G. and Craighero, L. (2004). The mirror-neuron system. *Annual review of neuroscience*, 27:169–192. 4

Rizzolatti, G., Fogassi, L., and Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature reviews. Neuroscience*, 2:661–670. 4

Robertson, J. V. G., Hoellinger, T., Lindberg, P. v., Bensmail, D., Hanneton, S., and Roby-Brami, A. (2009). Effect of auditory feedback differs according to side of hemiparesis: a comparative pilot study. *Journal of neuroengineering and rehabilitation*, 6:45. 4, 13, 41, 44, 134, 137

Roby-Brami, A., Feydy, A., Combeaud, M., Biryukova, E. V., Bussel, B., and Levin, M. F. (2003). Motor compensation and recovery for reaching in stroke patients. *Acta neurologica Scandinavica*, 107:369–381. 157, 169

Roby-Brami, A., Van Zandt-Escobar, A., Jarrassé, N., Robertson, J. V. G., Schnell, N., Boyer, E. O., Rasamimanana, N., Hanneton, S., and Bevilacqua, F. (2014). Toward the use of augmented auditory feedback for the rehabilitation of arm movements in stroke patients. *Annals of Physical and Rehabilitation Medicine*, 57:e4–e5. 157

Rocchesso, D., Polotti, P., and Delle Monache, S. (2009). Designing continuous sonic interaction. *International Journal of Design*, 3(3):13–25. 31

Rochesso, D. (1997). *Auditory Information Design.* PhD thesis, The Australian National University. 31

Rochesso, D. (2011). *Explorations in Sonic Interaction Design.* Logos Verlag, Berling, Germany. 31

Rodger, M. W. M. and Craig, C. M. (2011). Timing movements to interval durations specified by discrete or continuous sounds. *Experimental Brain Research*, 214(3):393–402. 25

Rodrigues, S. T., Vickers, J. N., and Williams, a. M. (2002). Head, eye and arm coordination in table tennis. *Journal of sports sciences*, 20(3):187–200. 162

Ronsse, R., Puttemans, V., Coxon, J. P., Goble, D. J., Wagemans, J., Wenderoth, N., and Swinnen, S. P. (2011). Motor learning with augmented feedback: modality-dependent behavioral and neural consequences. *Cerebral cortex (New York, N.Y. : 1991)*, 21(6):1283–94. 29, 89, 98, 130

Rosati, G., Oscari, F., Reinkensmeyer, D. J., Secoli, R., Avanzini, F., Spagnol, S., and Masiero, S. (2011). Improving robotics for neurorehabilitation: enhancing engagement, performance, and learning with auditory feedback. In *2011 IEEE International Conference on Rehabilitation Robotics (ICORR)*, pages 1–6. 48, 137

Rosati, G., Oscari, F., Spagnol, S., Avanzini, F., and Masiero, S. (2012). Effect of task-related continuous auditory feedback during learning of tracking motion exercises. *Journal of neuroengineering and rehabilitation*, 9(1):79. 27, 63, 89, 99, 136

Rose, F. D., Brooks, B. M., and Rizzo, A. a. (2005). Virtual reality in brain damage rehabilitation: review. *Cyberpsychology & behavior : the impact of the Internet, multimedia and virtual reality on behavior and society*, 8(3):241–62; discussion 263–71. 84

Rosenkranz, K. and Rothwell, J. C. (2012). Modulation of proprioceptive integration in the motor cortex shapes human motor learning. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 32(26):9000–6. 100

Rusdorf, S. and Brunnett, G. (2005). Real time tracking of high speed movements in the context of a table tennis application. *Proceedings of the ACM symposium on Virtual reality software and technology - VRST '05*, page 192. 162

Salmoni, A. W., Schmidt, R. A., and Walter, C. B. (1984). Knowledge of Results and Motor Learning : A Review and Critical Reappraisal. *Psychological Bulletin*, 95(3):355–386. 28, 88

Sarlat, L., Warusfel, O., and Viaud-Delmon, I. (2006). Ventriloquism aftereffects occur in the rear hemisphere. *Neuroscience letters*, 404(3):324–329. 54

Savary, M., Schwarz, D., Pellerin, D., Massin, F., Jacquemin, C., and Cahen, R. (2013). Dirty tangible interfaces. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems on - CHI EA '13*, page 2991. 32

Schaffert, N., Mattes, K., and Effenberg, A. O. (2011). Examining effects of acoustic feedback on perception and modification of movement patterns in on-water rowing training. *Proceedings of the 6th Audio Mostly Conference on A Conference on Interaction with Sound - AM '11*, pages 122–129. 40

Scheef, L., Boecker, H., Daamen, M., Fehse, U., Landsberg, M. W., Granath, D.-O., Mechling, H., and Effenberg, A. O. (2009). Multimodal motion processing in area V5/MT: evidence from an artificial class of audio-visual events. *Brain Research*, 1252:94–104. 23

Schmitz, G. and Bock, O. (2014). A comparison of sensorimotor adaptation in the visual and in the auditory modality. *PloS one*, 9(9):e107834. 28

Schmitz, G. and Effenberg, A. O. (2012). Perceptual effects of auditory information about own and other movements. *Proceedings of the 18th International Conference on Auditory Display*, pages 89–94. 23, 100

Schmitz, G., Kroeger, D., Effenberg, A. O., and Moritzwinkel, A. (2014). A mobile sonification system for stroke rehabilitation. In *Proceedings of the 20th International Conference on Auditory Display ICAD-2014*. 45

Schmitz, G., Mohammadi, B., Hammer, A., Heldmann, M., Samii, A., Münte, T. F., and Effenberg, A. O. (2013). Observation of sonified movements engages a basal ganglia frontocortical network. *BMC neuroscience*, 14:32. 23

Schnell, N. (2013). *Playing (with) Sound Of the Animation of Digitized Sounds and their Reenactment by Playful Scenarios in the Design of Interactive Audio Applications*. PhD thesis, University of Music and Performing Arts, Graz, Austria. 19

Schnell, N., Röbel, A., Schwarz, D., Peeters, G., and Borghesi, R. (2009). MuBu and Friends – Assembling Tools for Content Based Real-Time Interactive Audio Processing in Max/MSP. In *Proceedings of the International Computer Music Conference (ICMC 2009)*, Montreal, Canada. 74, 86

Scholz, D. S., Wu, L., Pirzer, J., Schneider, J., Rollnik, J. D., Groß bach, M., and Altenmüller, E. O. (2014). Sonification as a possible stroke rehabilitation strategy. *Frontiers in neuroscience*, 8(October):332. 45

Schwarz, D. (2011). State of the art in sound texture synthesis. In Proc. Digital Audio Effects (DAFx) (pp. 221-231). In *Proceedings of Digital Audio Effects (DAFx)*, pages 221–231. 22

Schwarz, D. (2012). The Sound Space as Musical Instrument: Playing Corpus-Based Concatenative Synthesis. In *New Interfaces for Musical Expression (NIME)*, pages 250–253. 86

Schwarz, D. and Schnell, N. (2010). Descriptor-based sound texture sampling. In *Proceedings of Sound and Music Computing conference*. 86

Secoli, R., Milot, M.-H., Rosati, G., and Reinkensmeyer, D. J. (2011). Effect of visual distraction and auditory feedback on patient effort during robot-assisted movement training after stroke. *Journal of neuroengineering and rehabilitation*, 8(1):21. 48

Seitz, A. R., Kim, R., and Shams, L. (2006). Sound facilitates visual learning. *Current biology : CB*, 16(14):1422–7. 4

Serafin, S. and Serafin, G. (2004). Sound design to enhance presence in photorealistic virtual reality. In *Proceedings of ICAD 04-TenthMeeting of the International Conference on Auditory Display*, Sydney, Australia. 31

Shams, L. and Seitz, A. R. (2008). Benefits of multisensory learning. *Trends in cognitive sciences*, 12(11):411–7. 4

Shing, C.-Y., Fung, C.-P., Chuang, T.-Y., Penn, I.-w., and Doong, J.-L. (2003). The study of auditory and haptic signals in a virtual reality-based hand rehabilitation system. *Robotica*, 21(02):211–218. 49

Shum, H. and Komura, T. (2005). Tracking the translational and rotational movement of the ball using high-speed camera movies. In *IEEE International Conference on Image Processing, ICIP 2005*, pages III – 1084–7. 162

Sigrist, R., Rauter, G., Marchal-Crespo, L., Riener, R., and Wolf, P. (2014). Sonification and haptic feedback in addition to visual feedback enhances complex motor task learning. *Experimental brain research.* 26, 40, 88, 152

Sigrist, R., Rauter, G., Riener, R., and Wolf, P. (2011). Self-Controlled Feedback for a Complex Motor Task. *BIO Web of Conferences*, 1(1). 26

Sigrist, R., Rauter, G., Riener, R., and Wolf, P. (2013). Augmented visual, auditory, haptic, and multimodal feedback in motor learning: a review. *Psychonomic bulletin & review*, 20(1):21–53. 29, 48, 88, 137

Sikström, E., de Götzen, A., and Serafin, S. (2014). The role of sound in the sensation of ownership of a pair of virtual wings in immersive VR. In *Proceedings of the 9th Audio Mostly: A Conference on Interaction With Sound.* 31

Smith, M. a., Ghazizadeh, A., and Shadmehr, R. (2006). Interacting adaptive processes with different timescales underlie short-term motor learning. *PLoS biology*, 4(6):e179. 16

Sorensen, V., Ingvaldsen, R. P., and Whiting, H. T. (2001). The application of co-ordination dynamics to the analysis of discrete movements using table-tennis as a paradigm skill. *Biological cybernetics*, 85(1):27–38. 164

Sosa, Y., Teder-Sälejärvi, W. A., and McCourt, M. E. (2010). Biases of spatial attention in vision and audition. *Brain and cognition*, 73(3):229–35. 64

Sowiński, J. and Dalla Bella, S. (2013). Poor synchronization to the beat may result from deficient auditory-motor mapping. *Neuropsychologia*, 51(10):1952–63. 34

Squire, L. R. and Kandel, E. R. (2000). *Memory: From mind to molecules.* Macmillan. 14

Steinbach, M. (1987). Proprioceptive knowledge of eye position. *Vision research*, 27(10):1737–1744. 109, 132

Stienstra, J., Overbeeke, K., and Wensveen, S. (2011). Embodying complexity through movement sonification. In *Proceedings of the 9th ACM SIGCHI Italian Chapter International Conference on Computer-Human Interaction Facing Complexity - CHItaly*, page 39, New York, New York, USA. ACM Press. 40

Subramanian, S. K., Massie, C. L., Malcolm, M. P., and Levin, M. F. (2010). Does provision of extrinsic feedback result in improved motor learning in the upper limb poststroke? A systematic review of the evidence. *Neurorehabilitation and neural repair*, 24(2):113–24. 13, 16, 42, 137

Sveistrup, H. (2004). Motor rehabilitation using virtual reality. *Journal of neuroengineering and rehabilitation*, 1(1):10. 49

Tajadura-Jiménez, A., Väljamäe, A., Toshima, I., Kimura, T., Tsakiris, M., and Kitagawa, N. (2012). Action sounds recalibrate perceived tactile distance. *Current biology : CB*, 22(13):R516–7. 4

Takeuchi, T. (1993). Auditory information in playing tennis. *Perceptual and motor skills*, 76(3 Pt 2):1323–8. 24, 136

Tammet, D. (2006). *Born on a blue day.* Free Press, New York, New York, USA. 69

Telford, K. (2006). Portable swing speed analyzer. US Patent App. 10/938,970. 162

Tepper, G. (2013). *ITTF Level 1 Coaching manual.* ITTF. 166

Terry, P. C. and Karageorghis, C. I. (2006). Psychophysical Effects of Music in Sport and Exercise : An Update on Theory , Research and Application. In *Australian & New Zealand Psychological Society*, pages 415–419. 40

Thaut, M. H. (2005). The future of music in therapy and medicine. *Annals of the New York Academy of Sciences*, 1060:303–8. 43

Thaut, M. H., Mcintosh, C., Rice, R., Miller, R. A., Rathbun, J., and Brault, J. M. (1996). Rhythmic Auditory Stimulation in Gait Training for Parkinson ' s Disease Patients. *Movement Disorders*, 11(2):193–200. 47

Thaut, M. H., McIntosh, G. C., and Rice, R. R. (1997). Rhythmic facilitation of gait training in hemiparetic stroke rehabilitation. *Journal of the neurological sciences*, 151(2):207–12. 47

Thoret, E., Aramaki, M., Kronland-Martinet, R., Velay, J.-L., and Ystad, S. l. (2012). Sonifying drawings : characterization of perceptual attributes of sounds produced by human gestures. In *Proceedings of the Acoustics 2012 Nantes Conference*, pages 1095–1100, Nantes, France. 25, 137

Thoret, E., Aramaki, M., Kronland-Martinet, R., Velay, J.-L., and Ystad, S. l. (2014). From Sound to Shape: Auditory Perception of Drawing Movements. *Journal of Experimental Psychology: Human Perception and Performance.* 25

Ticini, L. F., Schütz-Bosbach, S., Weiss, C., Casile, A., and Waszak, F. (2012). When Sounds Become Actions: Higher-order Representation of Newly Learned Action Sounds in the Human Motor System. *Journal of Cognitive Neuroscience*, 24:464–474. 4

Timmermans, H., Wolferen, G. V., Newland, P., and Kunath, S. (2004). MEDIATE : Key Sonic Developments in an Interactive Installation for Children with Autism. In *Proceedings of Int'l Computer Music Conference.* 84

Todorov, E. (2004). Optimality principles in sensorimotor control. *Nature neuroscience*, 7(9):907–15. 15

Todorov, E., Shadmehr, R., and Bizzi, E. (1997). Augmented Feedback Presented in a Virtual Environment Accelerates Learning of a Difficult Motor Task. *Journal of motor behavior*, 29(2):147–158. 41, 166

Tong, J., Stevenson, S. B., and Bedell, H. E. (2008). Signals of eye-muscle proprioception modulate perceived motion smear. *Journal of vision*, 8(14):7.1–6. 132

Uno, Y., Kawato, M., and Suzuki, R. (1989). Formation and control of optimal trajectories in human multijoint arm movements: Minimum torque-change model. *Biological Cybernetics*, 61:89–101. 15

van Vliet, P. M. and Wulf, G. (2006). Extrinsic feedback for motor learning after stroke: what is the evidence? *Disability and rehabilitation*, 28(13-14):831–40. 17, 28, 41, 43, 46, 88

van Vugt, F. T. (2013). *Sounds On Time.* PhD thesis, Université de Lyon. 26, 34, 35

van Wijck, F., Knox, D., Dodds, C., Cassidy, G., Alexander, G., and Macdonald, R. (2012). Making music after stroke: using musical activities to enhance arm function. *Annals of the New York Academy of Sciences*, 1252(1):305–11. 47

Varela, F. J. (1989). *Invitation aux sciences cognitives.* Seuil. 18, 19

Varela, F. J., Thompson, E., and Rosch, E. (1991). *The embodied mind.* MIT Press. 19

Vertegaal, R. and Eaglestone, B. (1996). Comparison of input devices in an ISEE direct timbre manipulation task. *Interacting with Computers*, 8(1):13–30. 38

Vinken, P. M., Kröger, D., Fehse, U., Schmitz, G., Brock, H., and Effenberg, A. O. (2013). Auditory Coding of Human Movement Kinematics. *Multisensory Research*, 26(6):533–552. 23, 24

Viviani, P. and Flash, T. (1995). Minimum-jerk, two-thirds power law, and isochrony: converging approaches to movement planning. *Journal of Experimental Psychology: Human Perception and Performance*, 21(1):32–53. 15

Vogt, K., Pirro, D., Kobenz, I., Höldrich, R., and Eckel, G. (2009). Physiosonic - Movement sonification as auditory feedback. In *6th International Symposium, CMMR/ICAD 2009*, pages 1–7, Copenhagen. 24, 100, 136, 152

Walker, B. N. and Kramer, G. (2004). Ecological Psychoacoustics and Auditory Displays: Hearing, Grouping, and Meaning Making. *Ecological psychoacoustics*, pages 150–175. 22

Walker, B. N. and Kramer, G. (2005). Mappings and Metaphors in Auditory Displays : An Experimental Assessment. *ACM Transactions on Applied Perception*, 2(4):407–412. 22

Wanderley, M. M. and Battier, M. (2000). *Trends in gestural control of music.* Ircam - Centre Pompidou. 36, 38

Wanderley, M. M. and Depalle, P. (2004). Gestural control of sound synthesis. In *Proceedings of the IEEE*, volume 92, pages 632–644. 35, 36

Wanderley, M. M. and Orio, N. (2002). Evaluation of input devices for musical expression: borrowing tools from HCI. *Computer Music Journal*, 26(3):62–76. 36, 37

Ward, J. and Meijer, P. (2010). Visual experiences in the blind induced by an auditory sensory substitution device. *Consciousness and cognition*, 19(1):492–500. 69, 70

Ward, R. and Morgan, M. J. (1978). Perceptual effect of pursuit eye movements in the absence of a target. *Nature*, 274(5667):158–159. 106, 107

Warren, J. E., Wise, R. J. S., and Warren, J. D. (2005). Sounds do-able: auditory-motor transformations and the posterior temporal plane. *Trends in neurosciences*, 28(12):636–43. 32

Wellner, M. and Zitzewitz, J. (2008). Evaluation of Visual and Auditory Feedback in Virtual Obstacle. *Presence*, 17(5):512–524. 31, 40

Wertheim, A. H. (2010). Motion perception during selfmotion: The direct versus inferential controversy revisited. *Behavioral and Brain Sciences*, 17(02):293. 109

Whitall, J., Waller, S. M., Silver, K. H. C., and Macko, R. F. (2000). Repetitive Bilateral Arm Training With Rhythmic Auditory Cueing Improves Motor Function in Chronic Hemiparetic Stroke. *Stroke*, 31(10):2390–2395. 47

Whitall, J., Waller, S. M., Sorkin, J. D., Forrester, L. W., Macko, R. F., Hanley, D. F., Goldberg, A. P., and Luft, A. (2011). Bilateral and unilateral arm training improve motor function through differing neuroplastic mechanisms: a single-blinded randomized controlled trial. *Neurorehabilitation and neural repair*, 25(2):118–29. 14

Wightman, F. L. and Kistler, D. J. (1989a). Headphone simulation of free-field listening. I: Stimulus synthesis. *The Journal of the Acoustical Society of America*, 85(2):858–867. 53

Wightman, F. L. and Kistler, D. J. (1989b). Headphone simulation of free-field listening. II: Psychophysical validation. *The Journal of the Acoustical Society of America*, 85(2):868–878. 53

Wing, A. M., Doumas, M., and Welchman, A. E. (2010). Combining multisensory temporal information for movement synchronisation. *Experimental Brain Research*, 200:277–282. 25

Winstein, C. J. and Schmidt, R. A. (1990). Reduced frequency of knowledge of results enhances motor skill learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(4):677–691. 28, 88

Wolf, P., Sigrist, R., Rauter, G., and Riener, R. (2011). Error Sonification of a Complex Motor Task. *BIO Web of Conferences*, 1:00098. 40, 134, 136

Wolpert, D. M. (1997). Computational approaches to motor control. *Trends in cognitive sciences*, 1(6):209–16. 20

Wolpert, D. M., Diedrichsen, J., and Flanagan, R. J. (2011). Principles of sensorimotor learning. *Nature Reviews Neuroscience*, 12(December):739–751. 17

Wolpert, D. M. and Flanagan, R. J. (2001). Motor prediction. *Current biology*, 11(18):R729–732. 15, 18

Wolpert, D. M. and Ghahramani, Z. (2000). Computational principles of movement neuroscience. *Nature neuroscience*, 3 Suppl(november):1212–1217. 12, 15, 100

Wolpert, D. M., Ghahramani, Z., and Jordan, M. I. (2005). An internal model for sensorimotor integration. *Science*, 269:1880–1882. 15, 18

Wolpert, D. M. and Kawato, M. (1998). Multiple paired forward and inverse models for motor control. *Neural networks : the official journal of the International Neural Network Society*, 11(7-8):1317–29. 15, 20

Wong, J. D., Kistemaker, D. a., Chin, A., and Gribble, P. L. (2012). Can proprioceptive training improve motor learning? *Journal of neurophysiology*, 108(12):3313–21. 100

Young, W. R., Rodger, M. W. M., and Craig, C. M. (2014). Auditory observation of stepping actions can cue both spatial and temporal components of gait in Parkinson's disease patients. *Neuropsychologia*, 57:140–153. 47, 50

Zatorre, R. J. (2003). Music and the Brain. *Annals of the New York Academy of Sciences*, 999(1):4–14. 32

Zatorre, R. J., Belin, P., and Penhune, V. B. (2002). Structure and function of auditory cortex: music and speech. *Trends in Cognitive Sciences*, 6(1):37–46. 35

Zatorre, R. J., Chen, J. L., and Penhune, V. B. (2007). When the brain plays music: auditory-motor interactions in music perception and production. *Nature Reviews Neuroscience*, 8(7):547–558. 32, 33, 34

Zatorre, R. J., Fields, R. D., and Johansen-Berg, H. (2012). Plasticity in gray and white: neuroimaging changes in brain structure during learning. *Nature neuroscience*, 15(4):528–36. 14

Zelic, G., Mottet, D., and Lagarde, J. (2011). Audio-tactile events can improve the interlimb coordination in Juggling. *BIO Web of Conferences*, 1:00102. 40