Analysis of the Embodied Cognition Process on People for Acquiring Music Skills through Games-based Learning

-Dissertation-

zur Erlangung des Grades
Doktor der Ingenieurwissenschaften (Dr.-Ing.)
der Universität Bremen

vorgelegt von

Katya Alejandra Alvarez Molina
aus Mexiko Stadt (Mexiko)

Bremen 2021

In memory of my father and Matilde

Gutachter:

Prof. Dr. Rainer Malaka

Prof. Dr. Christian Gütl

Tag der mündlichen Prüfung: Dezember 17, 2021

Abstract

Music has been present in human life, creating a social connectedness, originating changes in persons' minds and moods, and producing different reactions in the human body. These reactions depend on personal experiences, the body sense, and physical stimuli (auditory, visual, or tactile). For instance, unconsciously, when people listen to a song, they can move their feet in synchrony with the rhythm. To acquire music abilities, the synchronization between body and stimuli is fundamental, either for rhythm or melody. Nowadays, technology supports the acquisition of music skills through the interaction between users and devices and opens the possibility of enhancing music engagement and cognitive, motor, affective, and social skills. To date, people have their first contact with music through technologies such as video games (e.g., Guitar Hero, Rock Band). One of the advantages of video games as a learning tool is the freedom of developers to add, modify, or suppress certain stimuli. Therefore, a video game can create different interactions, movements, and sounds where the music guides the player to take a specific action. Understanding the relationship between music and video games will help design effective and efficient games for music learning purposes. However, what is less clear is what effects interactivity, movement, and sounds cause in music perception due to the stimuli emitted by video games. Therefore, there is a lack of connection between the design of video games and the embodied music cognition, where the cognitive, sensorimotor, social dispositions, and capabilities of human beings are sometimes not considered.

The goal and leading research questions for this thesis are to explore how the different stimuli emitted by a music video game change the players' performance. Moreover, how game-elements of a video game can improve music learning and user experience. This thesis seeks to understand the body reaction of players while playing music video games using different stimuli

and analyze the game elements and mechanics for music learning. Therefore, two music video games are presented, one game based on rhythms and one game based on pitch recognition. The thesis presents two case studies using the rhythm game to analyze players' reaction times and user experience with auditory, visual, or tactile stimuli. A case study compares the user experience between players using a video game and players using a web application for the pitch game.

The results show that players' performance is based on the game-elements and not on the stimuli; however, auditory stimuli enhance their performance over other stimuli. Moreover, a possible learning effect was observed after some trials (lives) and was found a fatigue effect during the gameplay. Using adequate game elements and mechanics can engage users in the activity at hand. In conclusion, human perception in video games is mainly focused on what the game shows visually; however, other stimuli impact the players' performance. Designing music game-based learning must find a balance between the game elements and mechanics and music perception. Consequently, players would acquire music skills during the gameplay.

Zusammenfassung

Musik war schon immer ein essenzieller Bestandteil der Gesellschaft. Sie schafft soziale Verbindungen, verursacht Veränderungen in den Köpfen sowie Stimmungen der Menschen und ruft unterschiedliche Reaktionen im menschlichen Körper hervor. Diese Reaktionen hängen von persönlichen Erfahrungen, den Sinnen und physischen Reizen (auditiv, visuell oder taktil) ab. Wenn Menschen beispielsweise unbewusst ein Lied hören, können sie ihre Füße synchron zum Rhythmus bewegen. Um musikalische Fähigkeiten zu erwerben, ist die Synchronisation zwischen Körper und Reizen von grundlegender Bedeutung, sei es für Rhythmus oder Melodie. Heutzutage unterstützt die Technologie den Erwerb von musikalischen Fähigkeiten durch die Interaktion zwischen Benutzern und Geräten und eröffnet die Möglichkeit, das Musikengagement und die kognitiven, motorischen, affektiven und sozialen Fähigkeiten zu verbessern. Bis heute haben Menschen ihren ersten Kontakt mit Musik über Technologien wie Videospiele (z. B. Guitar Hero, Rock Band). Einer der Vorteile von Videospielen als Lernwerkzeug ist die Freiheit der Entwickler, bestimmte Reize hinzuzufügen, zu ändern oder zu unterdrücken. Daher kann ein Videospiel verschiedene Interaktionen, Bewegungen und Geräusche erzeugen, bei denen die Musik den Spieler zu einer bestimmten Aktion führt. Das Verständnis der Beziehung zwischen Musik und Videospielen wird helfen, effektive und effiziente Spiele zum Musiklernen zu entwickeln. Weniger klar ist jedoch, welche Effekte Interaktivität, Bewegung und Geräusche auf die Musikwahrnehmung durch die Reize von Videospielen verursachen. Daher fehlt es an Verbindung zwischen dem Design von Videospielen und der verkörperten Musikkognition, wobei die kognitiven, sensomotorischen, sozialen Dispositionen und Fähigkeiten des Menschen manchmal nicht berücksichtigt werden.

Die leitenden Forschungsfragen dieser Arbeit untersuchen, wie die verschiedenen Reize, die von einem Musikvideospiel ausgehen, die Leistung

der Spieler verändern. Darüber hinaus wird untersucht, wie Spielelemente eines Videospiels das Musiklernen und die Benutzererfahrung verbessern können. Diese Arbeit versucht, die Körperreaktion von Spielern beim Spielen von Musikvideospielen unter Verwendung verschiedener Reize zu verstehen und die Spielelemente und Mechanismen für das Musiklernen zu analysieren. Daher werden zwei Musikvideospiele präsentiert, ein Spiel basierend auf Rhythmen, und ein Spiel basierend auf Tonhöhenerkennung. Die Arbeit präsentiert zwei Fallstudien, in denen das Rhythmusspiel verwendet wird, um die Reaktionszeiten und die Benutzererfahrung von Spielern mit akustischen, visuellen oder taktilen Reizen zu analysieren. Eine Fallstudie vergleicht die Benutzererfahrung zwischen Spielern, die ein Videospiel verwenden, und Spielern, die eine Webanwendung für das Pitch-Spiel verwenden.

Die Ergebnisse zeigen, dass die Leistungen der Spieler auf den Spielelementen basiert, und nicht auf den Reizen; Hörreize verbessern jedoch ihre Leistung gegenüber anderen Reizen. Darüber hinaus wurde nach einigen Versuchen (Leben) ein möglicher Lerneffekt beobachtet und ein Ermüdungseffekt während des Spiels festgestellt. Die Verwendung geeigneter Spielelemente und -mechaniken kann die Benutzer in die jeweilige Aktivität einbeziehen. Zusammenfassend lässt sich sagen, dass sich die menschliche Wahrnehmung in Videospielen hauptsächlich auf das konzentriert, was das Spiel visuell zeigt; andere Reize wirken sich jedoch auf die Leistung der Spieler aus. Bei der Gestaltung von musikspielbasiertem Lernen muss ein Gleichgewicht zwischen den Spielelementen und der Mechanik sowie der Musikwahrnehmung gefunden werden. Auf diese Weise würden die Spieler während des Spiels musikalische Fähigkeiten erwerben.

Contents

Chapter No. 1

Introduction

Is music so important? Music (rhythm and pitch) existed even before language; it was one of the first ways of communication between humans. However, music as we know it, with rhythm and melody, could not exist without the human body. The human body is responsible for most musical creations; each movement of the body can produce a sound by performing an instrument or even just by the move of our vocal cords. Music can exist even in our minds. According to our music experience, music can refer us to episodes in our life, moments and experience. Music can also generate emotions and feelings that produce an involvement with it. However, how can the human body and music be related? How does the interaction between the human body and music work to produce such feelings? The starting point to understand this interaction is to think about music from a different perspective. Music must be appreciated not only as the physical definition of energy, but also as the music-mind that exists in the brain of the people, and which is different from the physical concept (Leman, 2008). The music-mind is the process of music in the brain that implies our experiences, perception, meaning, sensory and motor systems. In this process, bodily movement plays a fundamental role in music perception (Leman, 2008; Maes et al., 2014; Leman, Marc and Maes, Pieter-Jan, 2015; Leman et al., 2018). For instance, when people listen to a song unconsciously they can start to move their feet in synchrony with the rhythm because their auditory-motor system is activated (Lahav et al., 2007). The cognitive process in music, such as learning, memory, and prediction, is related to the corporal interaction of humans, called embodied music cognition (Leman, Marc and Maes, Pieter-Jan, 2015; Leman et al., 2018). Moreover, bodily interactions of humans can be extended artificially through technology, allowing experimentation of new forms of mediation (Collins, 2011, 2013; Leman, Marc and Maes, Pieter-Jan, 2015; Leman, 2008; Leman et al., 2018). Thereby, humans use these mediation technologies, where the tool provides the interaction between the body and music. Mediation technologies open the possibility of enhancing the music engagement as well as the cognitive, motor, affective and social skills. To date, a broad gamut of technological tools have been used to study the embodied music cognition. Moreover, in many cases, people have their first contact with music through technologies such as video games (e.g., Guitar Hero, Rock Band) (Missingham, 2007).

One of the advantages of video games as mediation technology is the freedom of developers to add, modify, or suppress certain stimuli to create different experiences with the users. Video games can capture actions and transfer them to appropriate sensory outcomes and, in turn, these sensory outcomes generate a reaction towards a specific motor or social goal (Leman et al., 2018). On the one hand, human beings tend to anthropomorphize objects, especially those which involve an interaction. For instance, in an interactive video game, human can anthropomorphize some elements of the game. However, the anthropomorphism of the character does not only depend on the motion, but also on the attribution of the cause or intention, or even the sound (Collins, 2011, 2013). The elements such as the interactivity, the movement, and the sound allow creating an empathic experience, where the music "guides" the player to take an action based on the music (Liebe, 2013). Therefore, embodied music cognition is present when the player has a sonic reaction that corresponds with an action of the body through the game. In this sense, the game becomes an extension of the body itself (Collins, 2011).

Understanding the relationship between music and video games is essential for designing effective and efficient music games to be used for music learning purposes. Nowadays, playing video games has become an essential part of the life of children and youth (Seel, 1997; Gee, 2007). Youth believe that the use of music video games can provide access to music education for more people, independently of the social stratum (Missingham, 2007). Music video games can motivate people and have an essential role in education. Even including interactive music video games in the curricula can result in an exciting and meaningful experience for young people and teachers (Gower and McDowall, 2012). For instance, games such as 'Guitar Hero,' can help students improve their musical skills and support the embodied competence of players. The constant use of the interface impacts on the players' muscles and creates muscle memory, which improves the music skills (Jenson and De Castell, 2009; Kylie et al., 2011; de Castell et al., 2014).

In addition, learning through video games depends on the association of the cognitive elements, such as sounds, images or body movements of the player, instead of the game's content itself (e.g., the story or scenes) (Kylie et al., 2011). To illustrate this idea, in a traditional music system people learn and read a music notation, at the same time they need to practice rhythm or metric skills. While in a music game system, people unconsciously acquire the rhythm or metric skills and its connection with the visual objects of the game. Thus, it will be easier for people to change to a traditional music system (Kylie et al., 2011).

Despite the growing interest of the scientific community on embodied music cognition, there is still a gap between the application of embodied cognition principles and music games.

1.1 Problem Definition and Research Questions

Thanks to technology, scientists and designers can explore and analyze the interaction between humans, music, and technology. Although there have been many important advances in the field of embodied music cognition, gaps and challenges remain. For instance, it is not clear what effects interactivity, movement, and sounds cause in music perception (Leman, 2008; Maes et al., 2014; Leman, Marc and Maes, Pieter-Jan, 2015; Leman et al., 2018). Moreover, music perception through technological devices (mediators) is sometimes complicated because the use of devices is not intuitive for humans. Therefore, the design of a mediation-device must create a feeling of non-mediation, which means that the devices do not interfere with the music perception, instead they must enhance it (Nijs et al., 2012a; Maes et al., 2018). Regarding music games, it is not clear what effect that they produce on music perception; furthermore, there is not a criterion around the design to achieve the music learning purposes (Collins, 2011; Kylie et al., 2011; Collins, 2013). Additionally, there is a lack of connection between the design of video games and the embodied music cognition, where the cognitive, sensorimotor, social dispositions and capabilities of human beings are sometimes not considered. Thereby, in this work, the following general research questions arise:

1. Music video games contain several stimuli that can be perceived by humans; therefore, how do different stimuli affect the interaction between players and games?

2. How do music-based video games enhance the subjective human experience of players with video games?

3. What are the basic requirements that designers need to consider to create an interactive music-based video game for music education to fulfill its role as a learning tool?

In order to answer these questions, this thesis reviews some of the work that has been done regarding music-based interactive systems. This review also presents some

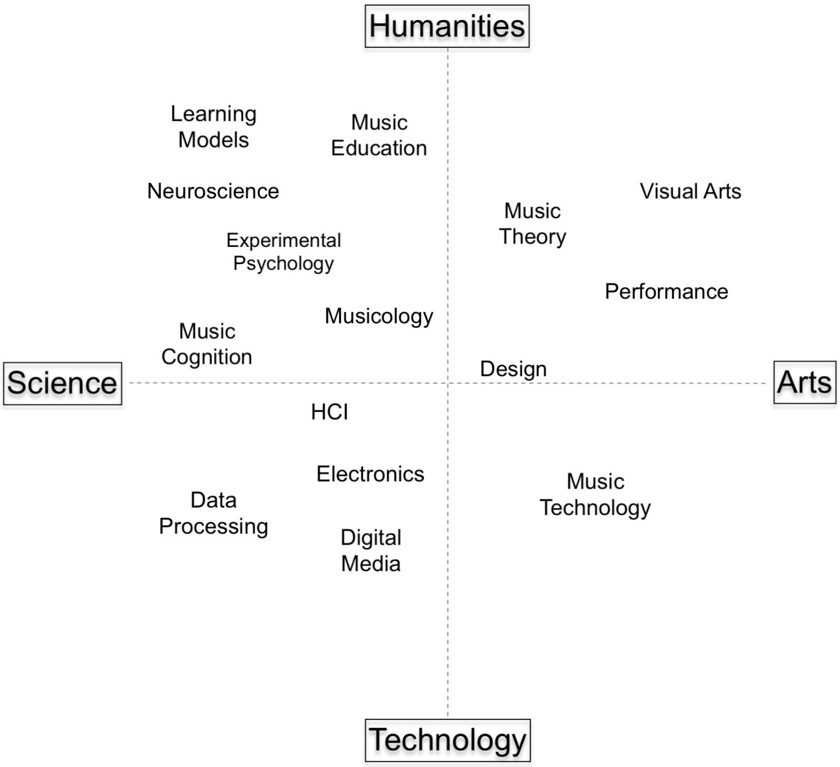


FIGURE 1.1: *The figure shows a map of the different areas that this thesis covers. Each axis represents a field of knowledge. The figure is based on a map presented by Jensenius (2007).*

concepts that are necessary to understand the relationship between systems, body, and music, and their interactions. Thus, these concepts cover different fields that are shown in Fig.1.1. The use of new technologies, moreover, offers new ways to assess and understand processes that underlie an engagement with music (Nijs et al., 2012a,b; Maes et al., 2018). Therefore, knowing what has been done and which aspects are related, offers the scientific community new starting points for research. In addition, studying the use of technology in the field of embodied cognition and music perception is important because technology has become somewhat intrinsic in human lives. For instance, some children and youth have their first contact with music through technologies such as video games (Jenson and De Castell, 2009; Gower and McDowall, 2012). Also, researchers must consider the potential that music video games have for music education; thus they need to assess, analyze, design, and develop these technologies to improve music skills (Kylie et al., 2011). In these processes, music perception and embodied cognition are important to achieve an optimal subjective experience and the acquisition of music skills Leman (2008). Hence, this thesis describes the developing process of two music video games, analyzes the strong and weak points of the design, and evaluates the different aspects of players using these games. In three cases studies, the performance of players is evaluated using different stimuli (auditory, visual, and tactile) in a video game environment. Consequently, the evaluation obtains qualitative and quantitative data such as reaction time, enjoyment, involvement, workload, and usability. Results are discussed in each study case followed by conclusions.

1.2 Outline of the Thesis

An overview of the thesis is presented in Fig. 1.2. The introduction, Chapter 1, describes the motivation for this work. Afterward, the first part of the chapter briefly mentions some related work and state of art about embodied music cognition theory from the perspective of using technology as a mediator and the interaction between music-based video games and players. Other ideas and concepts included in this chapter are the use of game-based learning and some theories of human perception. The second part of the chapter introduces the problem and the research questions, and how this work tries to solve them. State of the art and related work of this topic are discussed in Chapter 2, which is divided into three main aspects. The first aspect gives a general perspective of the role that embodiment cognition has in the relationship between music and humans and how it can be used to design interactive interfaces. The second aspect explores the use of music-based interactive systems, some game design theories, and players' game experiences. The third aspect presents a review of the use of these systems to enhance music skills. Subsequently, this part discusses game-based learning and how these topics can converge in order to find a balance for future designs of music-based interacting systems.

Chapters 3 and Chapter 4 present some prototypes along with case studies to illustrate the concepts generated in Chapter 2.

In order to answer some of the aforementioned research questions, Chapter 3 presents the design and implementation of a music video game. The prototype is based on rhythm sequences, where players follow the sequences. To determine the effect that different stimuli have on the interaction between players and the rhythm in the video game, the prototype was adapted to different stimuli. Moreover, this prototype also contributes to the question regarding the subjective experience of players using music

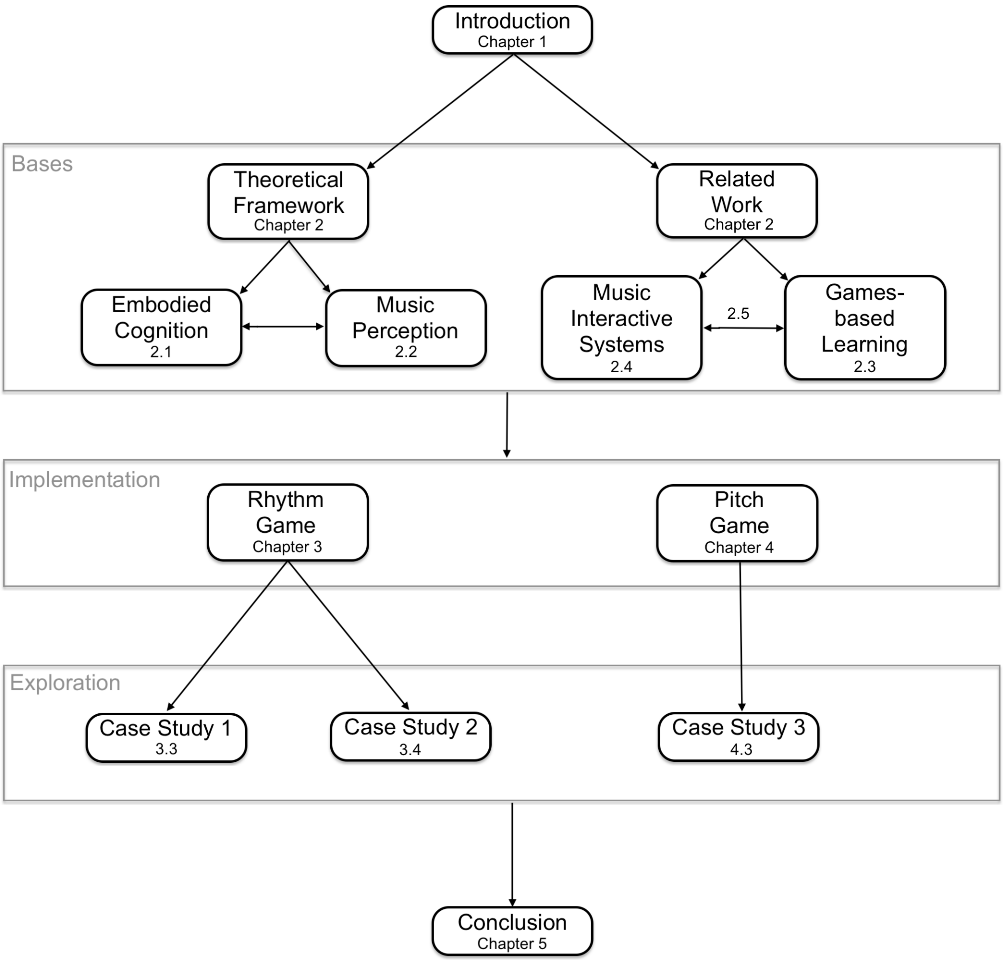


FIGURE 1.2: *The figure displays the general outline of the thesis, which is divided into three main parts. Also, the images shows the connection paths between each of the items described in this thesis.*

video games. For this purpose, the prototype presents rhythm sequences using auditory, visual, and tactile stimuli. Furthermore, Chapter 3 describes two case studies.

The first part of the chapter presents a case study of reaction times of players through visual and auditory stimuli (sequences), both with and without visual stimuli (game-elements) in the game. The study measures the workload of players during the game. Therefore, the hypotheses for this case study are:

- H1: Players have a better reaction time using auditory stimuli even in the absence of supportive visual game-elements.
- H2: Musicians have a better reaction time than non-musicians using auditory or visual stimuli.
- H3: The workload for participants using visual stimuli is higher than participants using auditory stimuli.

In the second part of the chapter, a case study compares the use of auditory vs. tactile stimuli by the reaction times and performance of the players. In addition, variables such as usability, enjoyment, and involvement are measured. The hypotheses are:

- H4: The performance of players using tactile stimuli is similar to the performance of players using auditory stimuli.
- H5: The absence of game-elements hinders the performance of players in both conditions.
- H6: The user experience is higher for participants using the auditory stimuli.

The end of the chapter discusses the use of different stimuli to improve rhythm skills based on the results.

The next chapter, Chapter 4, contributes to the research questions by the design and analysis of a music video game focused on pitch recognition. The first part of this chapter presents a brief review of video games to improve pitch recognition and describes the design process of the game. The prototype is based on auditory stimuli and a motor recognition mapping. Afterward, the chapter describes a case study where players evaluate their subjective experience using a prototype or web application. Moreover, this chapter considers an analysis of the learning tools used to design the prototype. The hypotheses for this chapter are:

- H7: Gamification of pitch recognition task is more attractive to youths than a no-gamified task.
- H8: Users learn a task such as pitch recognition by the use of a video game.

The discussion and conclusion chapter covers all the ideas presented in this research. The contributions of this thesis in the fields of Human-Computer Interaction and Musicology are described in this chapter. This chapter also contains some future work on the designs of music-based interactive systems from a broader panorama.

Chapter No. 2

State of the Art and Related Work

This chapter gathers some theories and concepts that support this work. As mentioned in Chapter 1, the approach of this thesis is multidisciplinary; therefore, this chapter starts with some concepts and cognitive theories, mainly regarding embodied cognition and its relationship with music and video games. The concepts presented in this part will be essential to understanding the approach of the subsequent chapters. Afterwards, Chapter 2 focuses on a review of theories and investigations regarding learning with video games and how cognition influences the performance of the players. Last part addresses the structure of music video games and their learning approach.

2.1 Cognition and the Movement of the Body

Cognition refers to human thoughts that are created by internal mental processes that involve interaction with the environment, physical actions, and communication with others (Tan, 2012). These internal mental processes include the attention, storage of memories, acquisition of knowledge and learning. In the course of human lives, persons modify their cognition according to their experiences in life. Experience with the world, body, and mind allows the reenactment of perceptual, motor, and introspective states, where the brain captures these states across modalities and integrates them into a multimodal representation stored in a memory. When some knowledge is required, the brain reactivates these representations by interacting with each other to simulate how the brain represents required knowledge (Barsalou, 2008; Engel et al., 2013). Over the last few decades, the human body has been a central part of cognitive sciences, creating a new research area referred to as *embodied cognition*. In this approach, sensorimotor networks are the basis of cognition, in particular, action-perception networks that are created to achieve a goal. In these networks, the environment plays a central role in shaping cognitive mechanisms (Barsalou, 2008; O'Regan and Noë, 2001).

Varela et al. (2017) described this approach as ”embodied action,” understanding cognition as the capacity of generating structure by action (enacting a world). Although Engel et al. (2013) state, also as a central premise, ”cognition is action” referring to an action-oriented paradigm (Clark and Chalmers, 1998; Clark, 1999; Schiavio and Altenmüller, 2015), they consider that action is not only a movement, but also intentional actions are grounded in sensorimotor coordination and contingencies, creating more complex forms of action.

An approach for analyzing the influence of body movements in the cognitive process is from the embodied, embedment, enactivism, and externalism (”4Es”) perspective (Menary, 2010; Schiavio and Altenmüller, 2015; Schiavio and van der Schyff, 2018; Gonzalez-Grandón and Froese, 2018). 4Es capture the interaction between body, brain, and environment in real-time. Schiavio and Altenmüller (2015) point out some key points:

- Cognition does not depend only on brain processes but results from structures widely distributed across the whole body of a living system (embodied mind).
- Cognition arises from interaction with the environment (social and physical); it is actively immersed in the world (embedded mind).

- Cognition can reach beyond the boundaries of the skull and skin, integrating resources both internal and external to the human (extended mind).

- Cognition consists of embedded and embodied forms of interactions between a self-organized living system and its environment. Through this dynamic interplay, the creature enacts (or brings forth), its domain of meaning (enacted mind).

Embodied cognition is considered as part of the situated cognition that takes place in the context of task-relevant inputs and outputs (Wilson, 2002; Gee, 2003). By definition, situated cognition involves interaction with the task that cognitive activity is at hand. Tasks imply an activity that is defined by time; therefore, situated cognition must deal with the constraints of real-time or run-time and must handle time pressure (Wilson, 2002). When time is an important factor, a human can confront a task in two ways. The first way is to rely on preloaded representations acquired through prior learning. The second way is to use the environment itself or external actions that the person performs to change his or her computational states (Wilson, 2002). The aim is to reduce workload. In the context of video games, playing *Tetris* (Pázhitnov, 1984) can be seen as a good tool to study human cognition because it is fast, repetitive, and requires split-second decisions of a perception and cognition. (Kirsh and Maglio, 1994). Players deal with decisions on how to orient and place each block by using actual rotation and translation movements, rather than by mentally computing a solution and then executing it (Kirsh and Maglio, 1994; Wilson, 2002).

During the performance of simple action, the motor system constructs a feed-forward simulation of the action to guide and correct the action (Barsalou, 2008). This simulation process, moreover, has high impact on social cognition through the observation and perception of the environment (stimuli); people represent these acts using simulations stored in their minds. Responsible for this ability are mirror neurons which discharge when an individual acts, as well as when the person observes a similar action performed by another person (Fadiga et al., 1995; Strafella and Paus, 2000; Gangitano et al., 2001; Buccino et al., 2004a; Rizzolatti, 2005; Cattaneo and Rizzolatti, 2009).

Behavioral studies have been investigating a paradigm about how movement observation could affect movement execution in stimulus-response compatibility. Studies of Brass et al. (2000) and Craighero et al. (2002) reinforced the theory of ideomotor actions, where creation of an image of sensory feedback related to a particular action is a fundamental step for executing an action (James, 1890; Greenwald, 1970; Weizsäcker, 1940; Buccino et al., 2004a). Nevertheless, this motor action depends on the observer's motor experience of a given action and the motor skills for that action. Motor action must be present in the motor repertoire of the observer to activate the mirror neuron system (Cattaneo and Rizzolatti, 2009).

These motor actions and activation of mirror neuron systems serve the purpose of understanding actions and intentions to imitate and generate empathy within the observed, e.g., feeling the same emotions that the others feel (Gallese et al., 1996; Gallese, 2005). In this role, mirror neutrons are capable of extracting two types of information: "What" the actor is doing and "Why" the actor is doing it; in consequence, it is possible to determine actor intentions (Wilson, 2002; Iacoboni et al., 2005; Rizzolatti, 2005). For instance, the hand grasps the cup (observation action "what") and the hand hand grasps the cup to drink ("why").

Nevertheless, the role of visual requirements to extract this information does not depend only on the association of the kinematics of the human body; instead, there is evidence that the mirror neuron system is sensitive to kinematics that do not match with the human kinematics at all (Gazzola et al., 2007). However, purely visual perception without the involvement of the motor system only describes the visual aspects of the actor and not about the intentions (Rizzolatti, 2005). Thus, it is imperative to put the observed action into a motor semantic network to understand the intention of the actor.

On one hand, it has been shown that visual stimuli are not the only stimuli that activates the mirror neuron system. Kohler et al. (2002) studied neurons in the F5 brain area that discharge when a monkey was performing a specific hand action and also when it hears the corresponding action-related sounds. Their findings showed a population of audiovisual mirror neurons discharging in this area during execution or observation of a specific action, but also when this action can only be heard. Thus, audiovisual mirror neurons plan/execute actions (motor condition) and recognize the action of others (sensory condition), and these neurons can evoke motor ideas even in the presence of only sound (Kohler et al., 2002; Keysers et al., 2003). In addition, Lahav et al. (2007) found a relationship between mirror neurons and musical sound, where sound-action mirror neurons were active while participants trained on a piece of piano music, indicating that the motor experience is critical for understanding sound actions.

The mirror neuron system is an essential part of imitation-based learning, where imitation is the translation of an action visually coded into an ideally a similar action by the observer Buccino et al. (2004b). Through imitation, a learner adapts and understand the actions of others because actions are represented visually, sonically, and gesturally in the mental process (Gebauer and Wulf, 1995; Kohler et al., 2002). For instance, when someone observes a musician performing, the brain recreates an internal model of the expressiveness at hand, in terms of the embodied experience of the person. Thus, the internal mental model activates the sensorimotor system to imitate the performer according to expressiveness. As a consequence, the music acquires a meaning of the emulated actions. This process to understand and mimic the gestures is unconscious and allows people to interpret the emotional inflections (Collins, 2011). In his mimetic hypothesis, Cox (2001) states how people understand body movement and human-made sounds in terms of their own previous experience of making similar movements and sounds. This hypothesis also explains why people continuously imitate the emotions and actions of external agents unconsciously. It is clear that the mirror neuron system is involved in the mimic process, generating a feeling of empathy with the originator, recreating visually and motorically the auditory stimuli in terms of intentionality, causality, and emotions. "Emotions generate movement, and movement generates emotions" (Bianchi-Berthouze et al., 2007).

Collins (2011) defines a term for the relationship between gestures and sound called Kinaesonic (Kinaesthetic + sonic). The term refers to an internal mapping between physical sounds and bodily movements. For instance, when a sound matches the action of our body, kinaesonic is congruent. Although physically, the body did not kinaesonically create a sound, the mimetic hypothesis and the mirror neuron system (through the mental and corporeal imitation) works to create a feeling that it did. Mimetics, in a video game context, allows players to engage in practices and learning experiences, procedures, and roles that more directly transform and transport the actor "in play" (de Castell et al., 2014). Body movement engages players with electronic devices, increasing not only the level of engagement, but also mediate feelings of presence in the digital world (Bianchi-Berthouze et al., 2007). When a body movement or action of a player matches with a sonic reaction in a video game, the sound is perceived as belonging to oneself (Collins, 2011) according to mimetic hypothesis. Therefore, game controllers (either with complex or basic designs) support the physical interaction and in addition to sound actions, can generate a feeling of empathy with the video game itself.

For instance, the controllers of *Guitar Hero* (Harmonix et al., 2005) or *Rock Band* (Harmonix and Pi Studios, 2007) have a design more specific to the game (e.g., guitar, drum, microphone), which allows the player to imitate the corresponding action with the visual cues displayed on the screen. However, a button-pushing controller, such as a Nintendo controller, also enables a mimetic engagement and impact on a player's thumb and finger muscles (Jenson and De Castell, 2009; de Castell et al., 2014). Furthermore, through a simple button, video games also acquire a meaning, forgetting about the physical device in order to concentrate on interacting with the events of the game (Griffin, 2005; Church, 2006).The button reduces complex action in a matter of choices (Griffin, 2005). Thus, new media requires controllers that afford the interaction and consider new metaphors and performative tasks; moreover, controllers must comprehend and specify the complexity of players' actions and embodied interactions.

When an external agent performs an action, and this action is well-known by an observer, the mirror neuron system of the observer is activated. As a consequence of the function of this system, a human can learn by imitation and generate empathy (Gallese et al., 1996; Gallese, 2005). The ability to understand the goals and intentions behind the actions of other individuals is a fundamental building block of social behavior. The absence of this ability is present in people with social isolating mental diseases such as autism (Iacoboni et al., 2005). In social interaction, feelings such as empathy, appear from the connection between feelings and emotions of others. Although empathy seems to be a subjective experience, it is a neurophysiological process that is created by an internal action-based process modulated by the mirror neuron system (Rizzolatti, 2005). Besides, the feeling of empathy appears not only when humans are involved; but also when objects or creatures generate this feeling. Humans tend to anthropomorphize objects or creatures that generate a feeling of empathy, trying to bring the object to life. This anthropomorphization is presented particularly in objects with which people usually interact. These objects provide feedback during actions, have human-like characteristics, or behave in a complex or intentional manner (Collins, 2011). The length of time that humans take to anthropomorphize simple objects is not long; a simple interaction with the object is enough (Morrison and Ziemke, 2005). Besides, the intentionality hypothesis of Tremoulet and Feldman (2000) states that when people can explain the intent of the movement of a single object, they will anthropomorphize the object. According to Cooley (1998) and Kohler et al. (2002), movement is not the only creator of anthropomorphization; sound activates the mirror neuron system, giving a sense of personality to inanimate objects. The findings previously pointed out have a direct relationship with video games, where players assign human-like physical characteristics through movement and sound. However, movement and sound sometimes are not enough to produce empathy. For instance, although films or animations contain movements and sounds, the phenomenon of anthropomorphization is not present (Collins, 2011). To produce empathy, it is important to have an interaction between the person and the animation or character. In a virtual world, the player can move the character from one position to another position using a game controller; thus, for the player, the character is alive. An internal map produces this sense of aliveness, whereby the player associates the movements of the character in the virtual world with the player movements in the real world (e.g., using a game controller) (Collins, 2011, 2013). Thus, the actions in the coordinate system of the external space become the actions that occur in the virtual game, creating a visuomotor mapping. Furthermore, an additional map (visuo-affective mapping) transforms the visual information about the external emotional states into similar emotional dispositions of oneself, influencing subjectivity, emotions, and empathy (Morrison and Ziemke, 2005)

2.2 Human Sensory System and Music Perception

2.2.1 Rhythm: Temporal Perception

Rhythm is present in daily life, in music, poetry, reading (Cummins et al., 1998; Cummins, 2009; Tierney and Kraus, 2015; Tierney et al., 2017; Woodruff Carr et al., 2017), even inside the human body like a heartbeat. Rhythm is essential for our body. The absence of it can be a symptom of some disease such as dyslexia (Overy et al., 2003; Overy, 2003; Huss et al., 2011), where persons cannot read fluently, or dyspraxia (Polatajko and Cantin, 2005; Roche et al., 2016), that refers to trouble with movement (Krampe, 2002). However, what is rhythm? The word rhythm comes from the Latin word "*rhythmus* (movement in time)" and the greek "*rhythmos* (measured flow or movement)" derived from "*rhein*" to flow. Therefore, rhythm can be defined as a pattern of temporal intervals in stimulus sequences (tones, clicks, vibrations, light) (Nguyen et al., 2018; Patel et al., 2005). Each one of these sensory patterns is called onset; the length between two onsets is called inter-onset interval (See Fig. 2.1).

Humans beings can perceive temporal patterns contained in the environment by cognitive and motor functions that originate beat perception and meter perception (Hannon et al., 2004). Beat perception is a psychologically perceived pulse that marks equally spaced points in time originated by rhythm; it is a mental event because it is not stimulus-driven, even though it usually arises in response to a rhythmic stimulus (Nguyen et al., 2018). On one hand, the meter is multiple levels of periodicity in a rhythmic structure in which some beats are perceived as more salient than others, as degrees of ascents or stress of individual beats (Hannon et al., 2004; Huang et al., 2013;

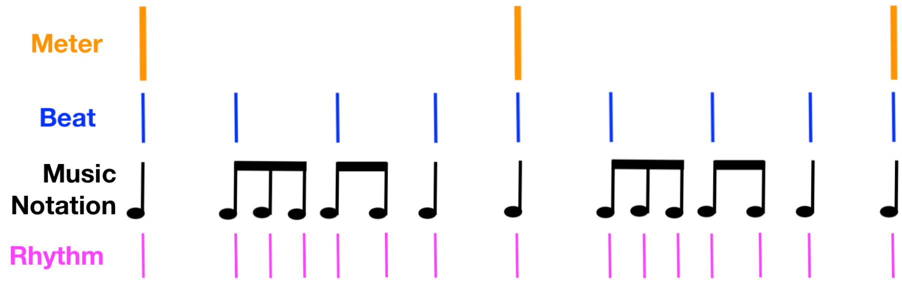


FIGURE 2.1: *In the rhythm (bottom), each vertical line represents the onset of the stimuli and the distribution of the sequence pattern. The vertical lines indicate the time between onsets. Musical symbols are the musical representation of the same rhythmic pattern. Whereas the beats, in blue lines, are the perceived events and the meter shows the metrical organization of the beats in rhythm (adapted from Nguyen et al. (2018)).*

Nguyen et al., 2018). It is the formal arrangement of the temporal patterns in music, into passages or sections, what Holst (1963) calls ”the time pattern of any song.” Depending on the meter, some metric structures can be more natural to feel the beat, in the case of a march (1 2 1 2) or a waltz (1 2 3 1 2 3), or difficult to feel in more complex meters (e.g., odd time signatures 7/8).

Understanding neural mechanisms that underlie processing of beat perception will help researchers learn how humans perceive and react to some stimuli and how humans can learn musical rhythms. Therefore, beat perception process in the brain has been studied from the perspective of behavioral investigations and neuroimaging studies. These studies have been focused mainly on sensorimotor synchronization where researchers have observed the ability of humans to move different body parts in synchrony with a pattern or beat using the sensory and motor system (Repp, 2005; Zatorre et al., 2007; Chen et al., 2008; Repp and Su, 2013; Novembre and Keller, 2018).

Findings using finger tapping and auditory stimuli in a synchronization task have found that the synchronization was less accurate when the beat is not clear than with a steady beat (Patel et al., 2005). Moreover, a comparison between different stimuli for the synchronization task showed that auditory stimuli dominate over visual stimuli in temporal perception (Repp and Penel, 2004; Patel et al., 2005; Kato and Konishi, 2006; Manning and Schutz, 2013; Hove et al., 2013a,b). This preference of the body to synchronize better with auditory stimuli is due to the active link between the auditory and motor systems of humans (Repp and Penel, 2002). Besides, studies suggest that visual stimuli dominate in spatial judgments because they offer more accurate sensory information (Repp and Penel, 2004; Hove et al., 2013b).

However, what happens when audio-visual integration is present in rhythm patterns? Experiments conducted by Repp and Penel (2004) and Kato and Konishi (2006) analyzed which stimuli attract more attention of participants or, in other words, which stimuli distract more in a synchronization task. On one hand, participants were following auditory rhythms, while flashes blinked in a different rhythm sequence as a distractor. On the other hand, participants were following rhythmic flashes, while auditory sequences could be heard in a different rhythm sequence as a distractor. Results show that participants felt more inclined to follow auditory sequences, even when the target stimuli to follow were the visual flashes; thus, participants tended to follow the auditory distractor sequences. Therefore, other sensory modalities, such as auditory, can manipulate visual perception in temporal tasks. For instance, when a single visual flash and multiple auditory beeps appear together, the single flash is incorrectly perceived as multiple flashes (Shams et al., 2000, 2002; Recanzone, 2003), e.g., a single flash accompanied by two auditory beeps is perceived as two different flashes (sound-induced illusory flashing). This phenomenon where auditory stimuli influence visual perception is often called ”temporal ventriloquism” (Burr et al., 2009) (Fig.2.2).

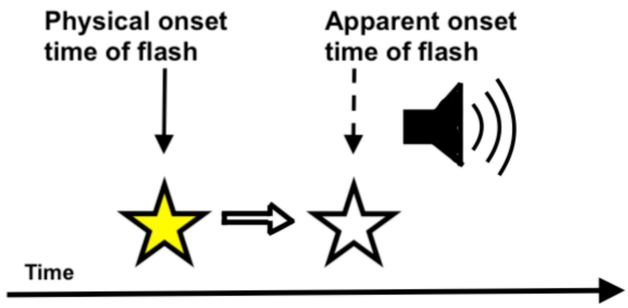


FIGURE 2.2: *Redrawn image from Chen and Vroomen (2013). Diagram of Temporal Ventriloquism. The perceived flash is shifted toward a sound that it is presented at a slightly different time than the flash.*

Nevertheless, although the synchronization task with auditory stimuli is better than with visual stimuli, studies found that visuomotor synchronization improved with a moving target (Bootsma and C. W. Van Wieringen, 1990; Tresilian, 1994; Hove et al., 2013b). For instance, in a target-distractor paradigma similar to Kato and Konishi (2006), Hove et al. (2013b) found, by using a bouncing ball, that distracting spatiotemporal visual rhythm can be as effective as a distracting auditory rhythm in a synchronization task. Furthermore, Hove et al. (2013b) found that the visuomotor synchronization was more stable with continuous moving targets than with discrete flashes; whereas auditory-motor synchronization was more stable with discrete beeps than continuous pitch-modulated sirens (Fig.2.3).

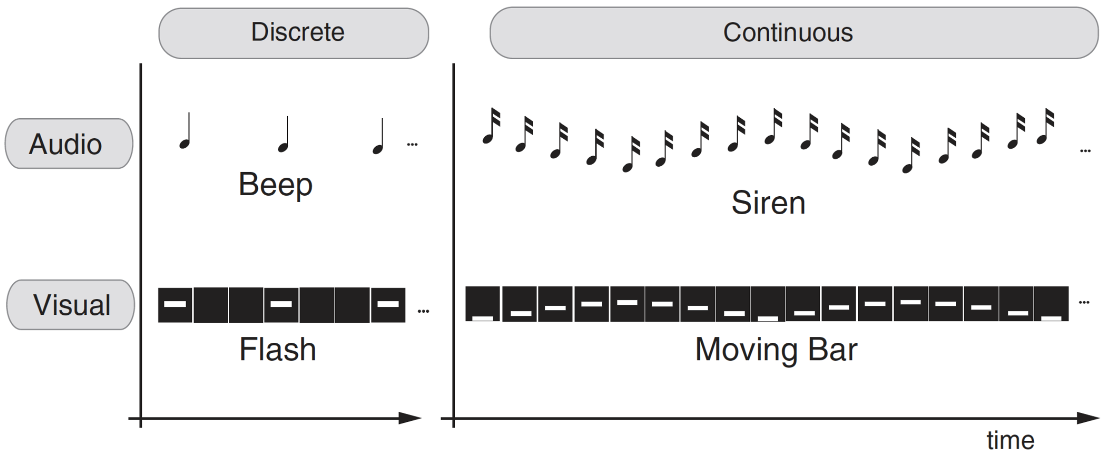


FIGURE 2.3: *Schematic of the flashes and beeps used in the study of Hove et al. (2013a).*

In addition, Iversen et al. (2015) found that a silent moving visual stimulus can drive rhythmic synchronization as accurately as an auditory metronome. To explain the dominance of one modality over another, researchers proposed three hypotheses (Welch and Warren, 1980):

- Modality precision hypothesis. Perception gives preference to the sensory modality best suited to the task at hand.
- Directed attention hypothesis. Visual stimuli do not have the strong alerting capacity of auditory stimuli; therefore, observers are predisposed to direct their at-

tention toward visual stimuli which, in turn, causes visual dominance (Posner et al., 1976).

- Modality appropriateness hypothesis. The human perceptual system is cognizant of the fact that vision is a more trustworthy modality for spatial localization than is audition and proprioception and, for this reason, it is more closely attended (Welch, 1999). Chen and Vroomen (2013) called this predicted effect temporal ventriloquism.

However, it has conclusively been shown that a synchronization task does not depend directly on modality; instead, it depends on the specific situation or, in other words, the nature of the stimuli (Wada et al., 2003; Hove et al., 2013a,b). For instance, synchronization with a visual metronome is less effective than with an acoustic metronome (Chen et al., 2002; Repp and Penel, 2002). However, researchers have observed that a visual rhythm can modulate temporal attention to visual targets (Schmidt et al., 2007; West et al., 2008; Escoffier et al., 2010; Donohue et al., 2010; Hove et al., 2013b) when musicians follow the beat of an orchestra conductor. To have a precise temporal synchronization between an action and the perception of different stimuli is crucial for interacting with different and complex situations that can be seen in human activities. For instance, precise visuomotor integration is needed to catch a ball, and audio-motor integration is necessary to synchronize movements with music (Hove et al., 2013a).

Therefore, researchers have explored the use of different stimuli to support rhythmic synchronization of tasks. These studies have been focused primarily on how humans perceive beat and meter using different stimuli such as auditory, visual, and haptic. According to the aforementioned literature, it is known that synchronization with purely visual flashes is more imprecise than with auditory stimuli (Chen et al., 2002; Repp and Penel, 2002), which indicate a strong relation between auditory and sensorimotor systems. It is unclear, however, whether beat perception is exclusive of auditory modality or whether humans can feel the beat and meter from other senses. Trying to find the difference in the synchronization between auditory and tactile stimuli, Brochard et al. (2008) evaluated the synchronization of participants following auditory or tactile beats. Strongly metered sequences (easy to follow beats) and weak metric sequences (complex to extract the temporal structure) composed the task (Fig.2.4).

FIGURE 2.4: *Redrawn image from Brochard et al. (2008) where the left column represents sequences with strong metric, meanwhile weak sequences are shown in the right column. Black bars represent an amplitude accent on that beat, while gray bars indicate regular beats. Empty squares represent silence.*

Results showed that the performance of participants using tactile stimuli was close to using auditory stimuli; therefore, they abstract the metric structure from the tactile rhythm sequences. Huang et al. (2013) found similar results to Brochard et al. (2008); however, they added that crossmodal sensory grouping between auditory and tactile senses enhance the meter perception. Thus, beat perception and meter are restricted not only to the auditory-motor system, but also to proprioceptive, somatosensory, vestibular, or tactile inputs (Brochard et al., 2008; Huang et al., 2013). All of these inputs can generate a sensation of "feeling" the music, but what it is happening is that mechanoreceptors cells (Pacinian) in the skin sense the pressure and vibrations as small as 100 angstroms (Talbot et al., 1968) and encode temporal patterns such as sound. Therefore, in the case of a tactile component, meter perception comes from the activation of these mechanoreceptors. On one hand, listening to a rhythm produces a physical movement (e.g., head moving from side to side) of the human body (Lahav et al., 2005; Trainor et al., 2009; Manning and Schutz, 2013) which implies the vestibular system. Trainor et al. (2009) substituted the head movement by an artificial stimulator located under each of the participants' ears and concluded that the device can influence the vestibular nerve and consequently, influence the interpretation of rhythm. Moreover, in the study of Manning and Schutz (2013), participants tapped along with a beat during a specific time, while during the same amount of time participants only listened without moving. Results showed that participants with the listen condition and without moving could not perceive the meter as well as participants under the moving condition.

To date, a variety of technological tools exist that improve the human body's capability to synchronize with stimuli. Video games provide an excellent example of the use of these tools in how users' rhythm skills are improved through the gamification process (Bégel et al., 2017, 2018). Moreover, games require accuracy in the reactions of the players to events or patterns that imply a proper synchronization between the body movements of the player and the stimuli of the video game. The players receive information from the video game via the senses to make an internal judgment, and they react with a movement of their body. Most video games transmit this information in the form of stimuli (visual or auditory).

Therefore, researchers have found that playing video games can result in an enhancement of various cognitive faculties that generalize beyond the original context (Latham et al., 2013; Boot, 2015). Cosper et al. (2009) found that the use of an interactive metronome in children with attention deficit and developmental coordination problems improved reaction time for making difficult visual choices and visuomotor control after training. Furthermore, some coordination problems originated by rhythm deficit have been treated by the use of rhythm video games finding high motivation in players (Castel et al., 2005; Bégel et al., 2017; Franceschini et al., 2017; Bégel et al., 2018).

Nevertheless, rhythm video games are based on repetitions and variations of stimuli. These types of games offer little freedom of expression to the player and contain restrictive rules on how the player is to react (Pichlmair and Kayali, 2007). Furthermore, the interaction between the player and the music is proactive, meaning players undertake a specific action when the stimuli appear (Liebe, 2013). Some examples of rhythm games in which players follow a sequence of visual and auditory patterns are

FIGURE 2.5: ***A)*** *Screenshot of the video game Elite Beat Agents (Video Games Museum, 2006) where players have to press the numbers in ascending order and in synchrony with the song of the game, e.g., number 1 appears in a determined beat of the song and players must press number 1. The player accumulates points for each correct action.* ***B)*** *Screenshot of the game Patapon (Lipke, 2017) displays a tribe that goes forward according to a rhythm of the song. The player sends commands created by the combination of four buttons (pata = square, pon = circle, chaka = triangle, don = cross) timed with the game's rhythm. Each sequence of each command refers to an action, e.g., attack, defend, retreat, and so on.* ***C)*** *Screenshot of Bit.Trip Runner (Reeves, 2011). This game is an autoscrolling platform game and jump-and-run dynamic, where rhythm is the main element.*

Patapon (Pyramid and SCE Japan Studio, 2007), *Rhythm Tengoku* (SPD, 2006), *Elite Beat Agents* (iNiS, 2006), or *Bit Trip Runner* (Gaijin Games, 2010) (Fig.2.5).

However, the visual or auditory elements of the game could be a critical factor for the performance of the player and for acquiring rhythm skills. When a rhythm game has more than the necessary audio or visual elements, the attention of the player is more focused on these elements than tapping the sequence. Thus, the synchronization task becomes difficult which affects the learning process. Moreover, developers have implemented the use of tactile stimuli in their design, where vibrotactile sensors enhance one's music experience (Yuan and Folmer, 2008; Seim et al., 2016). One of the aims of these designs was to introduce music to deaf people, where the designs obtained good results regarding the effectiveness and acceptance of people (Nanayakkara et al., 2009; Baijal et al., 2012; Jack et al., 2015). Another approach to the use of tactile sensors is in-game console controllers, where the use of vibrotactile sensors improves the game experience, such as game *Mario Kart* (Nintendo EAD and Namco/Namco Bandai, 1992). However, the use of vibrotactile sensors in music video games, in particular, rhythm games, is not clear regarding the behavior of players using these sensors.

2.2.2 Mapping the Pitch: Activation in the Brain

As shown in the prevous sections, for musicians as well as non-musicians, music can evoke motor action because a strong link between perception and action exists. For musicians, playing an instrument requires the integration of auditory and motor processes where the stimulation of one modality can lead the activation of another modality (Kohler et al., 2002). Thus, musicians correlate modalities by creating a map of sound ″musical ear″ and its respective movement of a limb ″musical hand″ in a so-called key-to-pitch paradigm (Bangert et al., 2001; Bangert and Altenmüller, 2003; Bangert et al., 2006). Several studies have found that a key-to-pitch map emerges in the brain (Fig.2.6); thus, activity has been observed in the sensorimotor cortex, anterior frontal, and temporal areas predominantly of the right cerebral hemisphere (Bangert and Altenmüller, 2003; Lahav et al., 2007). These areas are related to real and imagined perception of melodic and harmonic pitch sequences, where the temporal pole is associated with the auditory process, the motor activity is connected to the dorsal frontal lobe, and ventrolateral or supraorbital part of the prefrontal cortex is related to the mapping (Bangert and Altenmüller, 2003; Gaab et al., 2003).

However, activation in these brain areas is restricted not only to the auditory-motor process, since a well-practiced piece of music can also trigger respective movement in musicians during passive listening (Haueisen and Knösche, 2001). Bangert and Altenmüller (2003) evaluated music beginners over five weeks (ten sessions) practicing a melody using their right-hand and following the note order, rhythmic timing, and loudness of the piano keystrokes. Participants first listen to the piano melody and, af-

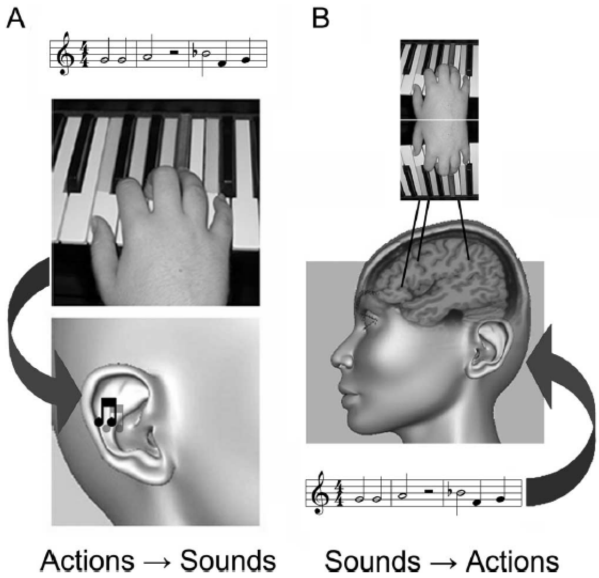


FIGURE 2.6: *Image taken from Lahav et al. (2007).* ***A)*** *Sound made by an action, the auditory stimulus is listened.* ***B)*** *The corresponding motor representation of the learned music is displayed.*

ter a pause, they tried to replicate the melody. Twenty melodies of five notes (C to G) composed the task, starting with melodies of only three notes and then increased the difficulty up to five notes. Rating categories consider pitch errors, timing errors, and dynamic errors. Participants were split into two groups to perform the task described above. One group, ”map group,” assigned the same finger to the same note, whereas the second group, ”no-map” were instructed to assign a different finger to different notes. The result was an auditory-sensorimotor activation in the first few minutes of playing which was reaffirmed within a few weeks. The map group learned the standard piano key-to-pitch map, and results showed activation in the brain area, contrary to the no-map group where activation was absent. Unlike the belief of some music educators that ear-to-hand ability is acquired after decades of practice, this study shows that this ability can be acquired within the first weeks of practice and can be the base of any instrumental skills. In a similar study, Lahav et al. (2007) found that in a recently learned melody, ”key-to-pitch” mapping is present in participants and activates the auditory-motor system network in the human brain. In their study, musically naive participants learned and practiced an original melody for five days. The trained melody consisted of four notes which was performed by participants who always used the same finger for the same note in a piano keyboard (action-listening condition).

After the fifth day, participants were scanned in functional magnetic resonance imaging (fMRI) while participants passively listened to a fragment from the trained melody. In a different passive listening task, participants listened to a new fragment of a melody with the same notes but in a different order. Afterwards, participants passively listened to a new fragment of a melody but with different notes. After each scan, participants carried out a behavioral control task which consisted of three heard piano notes and pressed a button with their left hand to indicate whether the notes had appeared in the preceding fragment of the melody. Results indicate that participants learned the piece gradually until an error-free performance of approximately 12 minutes in the last session. The test showed that subjects developed (albeit not to perfection) a ”pitch-to-key” mapping, which means the ability to recognize the pitches and identify them in real-time on the piano keyboard. Participants improved 53% from pre-test to post-test on the pitch-recognition-production test.

Studies of Bangert and Altenmüller (2003) and Lahav et al. (2007) agree with findings of Baumann et al. (2005) and Kohler et al. (2002) regarding the acquisition of key-to-pitch mapping ability and its relationship with the auditory and sensorimotor system. Furthermore, these findings suggest that this ability is acquired at the beginning of the practice and settles over time. However, ”key-to-pitch” mapping is not the only way to activate the motor system while playing an instrument. During sight-reading, musicians translate the visual-symbolic representation of a note into a motor command (Jäncke, 2006). For instance, when a musician sees the image of note F in a music sheet, the musician strikes the key for note F on the keyboard automatically. The distance between the eye (fixation of a note) and hand position (tapping the corresponding key) is called eye-hand span (EHS) which is the anticipatory time span (length of time between the fixation and performance of a note in s or ms) or by notes (number of notes between hand and eye fixation) (Truitt et al., 1997; Rosemann et al., 2016). Novice musicians translate each note into its corresponding movement of a limb, unlike expert musicians who look at a series of notes and merge that immediate information with knowledge held in long-term memory (Wristen, 2005; Stenberg and Cross, 2019). Thus, many musicians develop the ability to improvise because they recode information in meaningful groups, reducing the workload of memory; meanwhile, the motor system processes the information (Rosemann et al., 2016).

On one hand, either key-to-pitch mapping or sight-reading in the early stages can produce muscle memory. Therefore, some methods for learning how to play an instrument start with the visual association of the note and its corresponding finger. For instance, Ramos Mejia (1947) in his method to learn violin started with hand exercises where the apprentice places the fingers in the correct position indicated by the music sheet. These exercises aim to train the muscles to use only the amount of energy necessary to press the string, establish the different grades of fingers distention and flexor.

During the last decades, a broad gamut of music interactive interfaces has been developed using virtual reality, augmented reality, PC or mobile games, and web applications. The aims of these interfaces are diverse, either to learn a melody in a piano keyboard, to identify a pitch, or to learn music theory. However, all these interactive systems have one common feature: they are based on the auditory-motor theory.

Some of the most popular interfaces, which have been evaluated by musicologists, computer scientists, and psychologists, are *Rock Band* (Harmonix and Pi Studios, 2007) and *Guitar Hero* (Harmonix et al., 2005). These systems contain a game controller to simulate a real instrument, such as a guitar, drum set, or keyboard, connected to a visual and audio interface that demostrate the game. Players press buttons and strum on the controller in synchrony with notes that scroll on-screen, where synchronized actions generate points. Researchers have found that while players are not going to learn an instrument using these types of video games, players can start to prepare their muscles for future learning of an instrument (Jenson and De Castell, 2009; Kylie et al., 2011; de Castell et al., 2014). Another advantage of the use of these games is the correlation that players generate between music symbols of the game and real music notation(Kylie

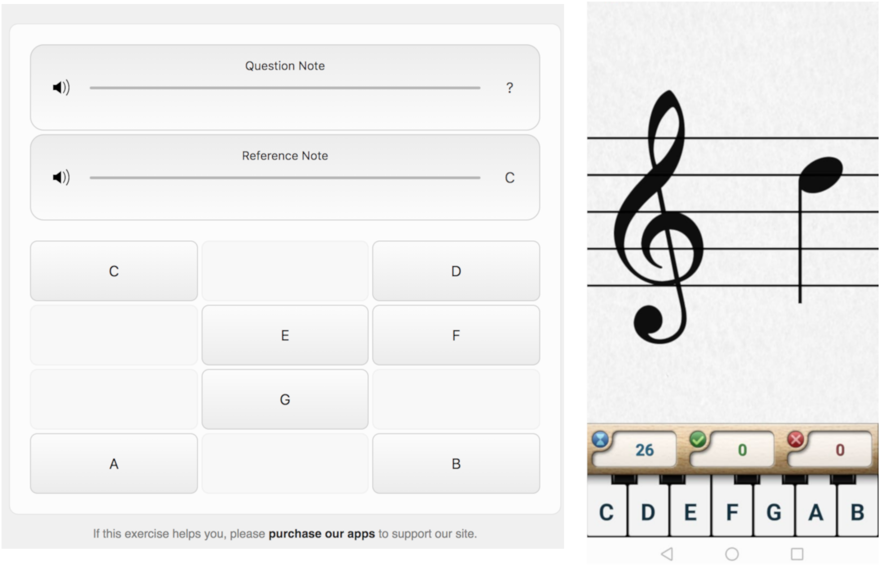


FIGURE 2.7: ***Left:*** *In-game screen of the web application musictheory.net (Musictheory.net, 2000) where a player listens to the pitch given by the game and needs to choose the correct name of the note that corresponds with the listening tone. The game can be configured by changing the range of the indexes, note names or instrument sound. The reference note helps the player construct an interval in his mind to identify the note easily.* ***Right:*** *In-game screen of the mobile application Music Tutor Sight Read Lite (VirtualCode.es, 2018). In reading notes modality, the application shows a note on a stave and also emits the sound of that note. Then, the player must touch over the correct note on the piano keys located on the bottom of the application. Three boxes also indicate (from left to right) a timer, correct answers, and incorrect answers.*

et al., 2011). The game uses its symbols to represent music, and the player transfers the meaning of these symbols into a real music notation system. For instance, when musicians are sight-reading, a similar process occurs. They see a music symbol on the staff[1], and they automatically move the corresponding finger of that symbol (Jäncke, 2006). Therefore, there is evidence that the use of interactive music interfaces support music learning when such interfaces use auditory-motor theory.

Nevertheless, some web or PC games for pitch recognition do not consider the auditory-motor theory in their designs. These applications (e.g., *Note Ear Training* (Musictheory.net, 2000), *Pitch Perfect Ear Training* (WholeTone-Games, 2018), *Music Tutor*

[1] Also spelled stave (UK), in the notation of Western music, five parallel horizontal lines that, with a clef, indicate the pitch of musical notes (The Editors of Encyclopaedia Britanica, y 15).

Sight Read (VirtualCode.es, 2018)) use a computer mouse or a touchpad to interact with the application (Fig.2.7). Therefore, there is no association between the sound and the corresponding body movement, which affects the creation of internal mental mapping. Consequently, players do not generate muscle memory which affects the the learning process. These applications are based on traditional music lessons and they are designed with simple designs (without game-elements or complex dynamics of interaction). Consequently, people are not engaged with these applications; thus, they desist engaging in the task.

On one hand, some studies have shown the benefit of using interactive dynamics on the design of applications, such as video games, to enhance the learning process (Baker et al., 2010; Juul, 2010; Boot, 2015; Hamari et al., 2016; Juul, 2003). These dynamics involve player experience, which considers the enjoyment and involvement of players with the game (Gajadhar et al., 2008). Player enjoyment indicates the amount of pleasure or displeasure in terms of positive affect, competence, challenge, frustration, and aggression. During player involvement, the player is wholly focused and interested in the game. Player involvement includes flow, immersion, competence, and challenge Gajadhar et al. (2008). Nevertheless, two states that have a more significant influence on the learning process are engagement and challenge (Schiefele, 2001; Cheng et al., 2015; Hamari et al., 2016).

2.3 Learning through Video Games

Playing video games has become an essential part of the lives of children and youth (Gee, 2007), and it allows them to be active participants rather than being only observers, as with television (Seel, 1997). To understand the scope of video games is important to define the word *game*. Based on previous definitions of games (See Appendix A A.1), Juul (2003) describes that a game is a combination of rules with negotiable consequences; it has variable and quantifiable outcomes with different values that the player tries to influence, and the player feels attached to these outcomes. Games deal with decisions, actions, and reactions, and these elements have clear consequences. Egenfeldt-Nielsen (2011) adapted the definition of Juul (2003) to video games which consider two main components for video games: verbs and substantives. The verbs are what players can do within the story and the environment. Whereas the substantives make up the story and the environment; they are only representations that make the game more immersive and with a purpose. Without a verb, however, a substantive would continue being only a representation (Fig.2.8).

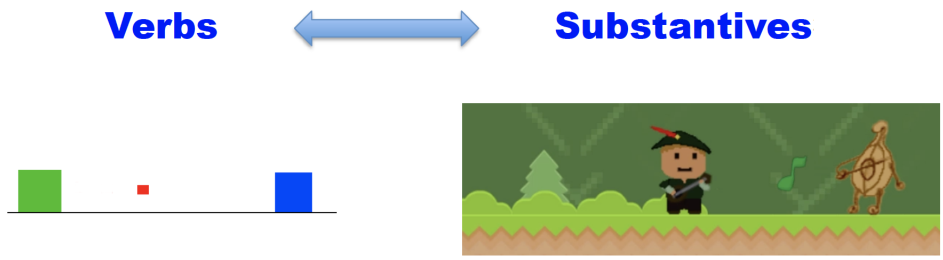


FIGURE 2.8: *Left side: A representation of verbs, an action. Green square launch a red bullet to the blue square. Right side: The same action is represented using the same number of elements but this time as substantives. Thus, verbs and substantives enrich the experience of the player.*

Although there is a wide variety of video games, several of them consider groups of people who are experts in video games, e.g., role-play games. Nevertheless, some video games create more inclusive games with simple designs, avoiding complex dynamics in comparison to other types of video games, which require more experience, such as role-games (Juul, 2010). These simple games are called casual games, and anyone can be a player. Casual games are a popular genre which gained popularity with the dissemination of smartphones. Some casual games can be interpreted as serious games, where the task component is more critical than visual design. Thus, they avoid having complex models to have less distortion of the information for players (Egenfeldt-Nielsen, 2011). These games are a recurrent genre in games whose purpose is learning.

Researchers suggest that designing serious games or game based learning depends not only on visual design but also on the player experience. Gajadhar et al. (2008) divides the subjective player experience in player involvement and player enjoyment.

Player involvement concerns flow, immersion, and engagement, where the player (in a state of concentration) is wholly focused and interested in the game. Furthermore, player enjoyment indicates the amount of pleasure or displeasure in terms of positive affect, competence, challenge, frustration, and aggression. For instance, positive affect measures the fun and enjoyment of gaming. In contrast, competence evaluates whether a player feels skilled in playing the game, and challenge measures the amount of effort that players believe they exert. Players are motivated and engaged when they feel competent during the game, where success is positively valued and failure is negatively valued. Nevertheless, competence focuses more on the perception of players regarding performance, where players can feel uncomfortable or insecure about their performance, originating a feeling of incompetence. This situation can cause the player to refrain from taking risks or new challenges, affecting the learning (Fullagar et al., 2013).

All these aspects of the experience with video games are an excellent opportunity to incorporate video games into the education field because they provide interactivity, user-centered design, and state-of-the-art computer technologies (Denis and Jouvelot, 2005). But the main reason for the use of video games in scholarly activities is motivation and engagement during the gameplay. For instance, understanding motivation is a key point in knowing why and how people take up learning a musical instrument and why they continue with the activity or quit. Motivation is described by the Self-Determination Theory (SDT) (Deci and Ryan, 2000), which includes three basic psychological needs for a person: relatedness, competence, and autonomy.

- Relatedness: feeling connected to others, a sense of belongingness with other persons or a community.
- Competence: feeling effective in one's behavior, experiencing opportunities to exercise and express one's capacities.
- Autonomy: being the perceived origin or source of one's behavior, individuals experience their behavior as an expression of self.

Researchers in SDT have studied the relationship between doing something that is interesting or pleasurable (intrinsic motivation) or doing something for another reason other than the task itself (extrinsic motivation) (Deci and Ryan, 2000; Evans, 2015). A motivated student cannot stop learning (intrinsic motivation) and activates efficient cognitive strategies for long-term memory. Satisfying SDT increases the chances of experiencing efficient learning (Denis and Jouvelot, 2005). Intrinsic motivation deals with fun factors, a reason why video games are useful. Denis and Jouvelot (2005) proposed an organization of the fun factors into two poles, pleasure and desire, and ludic tension (Harter, 2002):

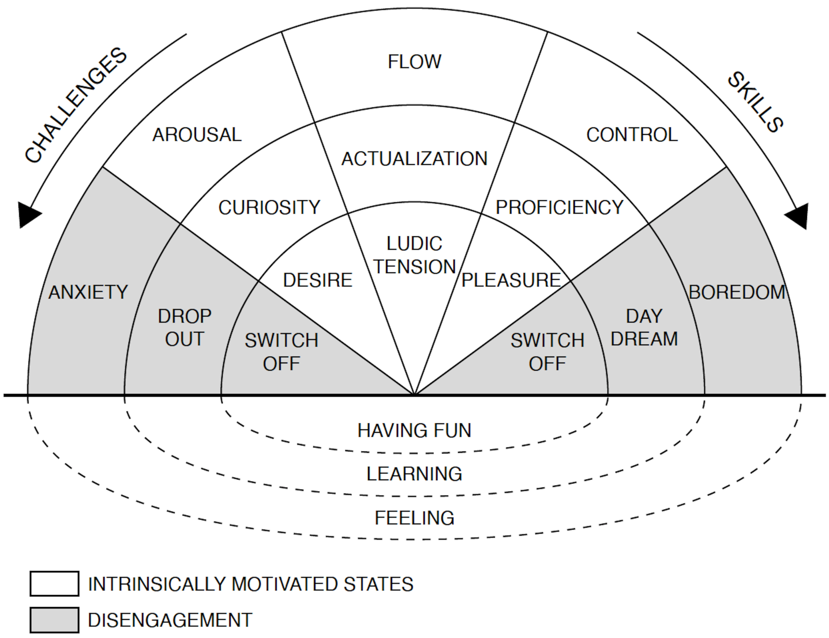


FIGURE 2.9: *Figure taken from Denis and Jouvelot (2005). Intrinsic motivation as a balance between challenges and skills.*

- ”Pleasure comes from fantasy (Malone, 1980), control (Malone and Lepper, 1987), power, creation, social interaction, immersion and comedy (Garneau, 2001), direct system response and experience of effectance (Klimmt, 2003), emotions or achievement of desire;” Denis and Jouvelot (2005).

- ”Desire comes from challenge and curiosity (Malone, 1980), problem solving and competition (Garneau, 2001), escapism and competence (Klimmt, 2003);” Denis and Jouvelot (2005).
- ”The ludic tension[2] comes from discovery (Garneau, 2001), conflict, suspense, and relief (Klimmt, 2003), learning (Koster, 2013), surprise, or narration.” Denis and Jouvelot (2005).

The organization (Fig.2.9) reflects the dynamic learning curve, where not only flow is related to intrinsic motivation but also arousal and control. These intrinsically motivated states can increase when a person participates in physical practices in any context.

Therefore, player experience in a video game has a close relationship with the learning experience.

However, some feelings affect learning more than others, such as flow, frustration, or boredom. Flow refers to a state of mind where an individual forgets all unpleasant aspects of life, leaving no room for distractions or outside information; attention is on relevant stimuli (Csikszentmihalyi, 1997; Csikszentmihalyi et al., 2014). Novak and Hoffman (1997) defined flow in terms of experiences (intrinsic enjoyment, loss of self-consciousness), structural properties (a seamless sequence of responses facilitated by interactivity with the computer and self-reinforcement), and antecedents (skill/challenge, balance, focused attention, and telepresence) (See Appendix for all definitions of Flow). Flow is characterized by full concentration on the task at hand, whereby the person loses track of time and self-consciousness in the sense of merging with one’s environment. The person has a sense of control: knows what to expect, how to deal with different situations, what needs to be done, and how to accomplish it. However, the task at hand in flow condition is a goal in and of itself, where the process is more important than outcomes of the activity (Strati et al., 2011; Fullagar et al., 2013; Csikszentmihalyi et al., 2014). Persons enjoy the process more than the goal; they do not pursue the activity for money, recognition, or power. This enjoyment does not

[2]Ludic tension describes the unstable whirlwind existing between pleasure and desire, chance and strategy, rules and freedom, reality and fiction (Harter, 2002).

mean that the person is "totally happy"; instead, the task can be physically or mentally demanding. Consequently, the person feels satisfied throughout the entire experience. Another necessary condition of flow is immediate and continual feedback on how well the person is progressing (Csikszentmihalyi, 1997; Strati et al., 2011; Fullagar et al., 2013; Csikszentmihalyi et al., 2014). Positive or negative feedback is essential to let the person know how well or bad the person is performing, so that the person can learn from this feedback and gain experience in the task. This feedback is fundamental, especially while learning a new skill.

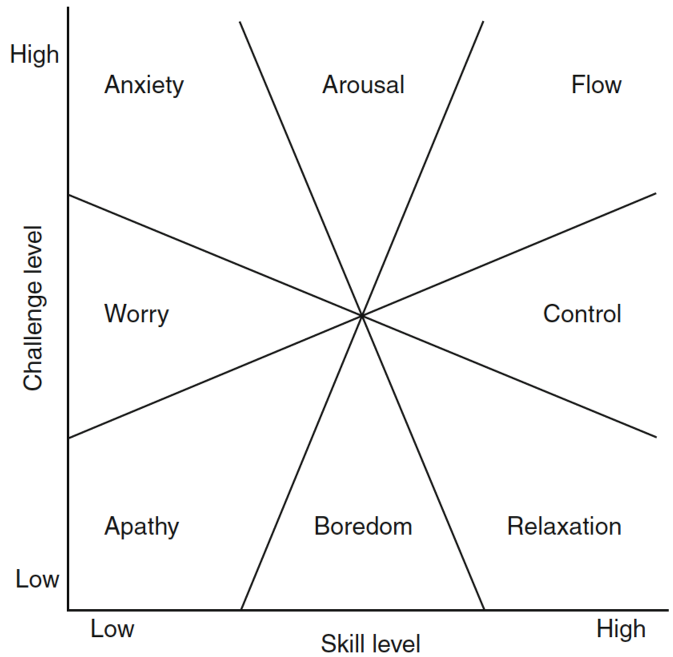


FIGURE 2.10: *Figure taken from Strati et al. (2011). Eight flow channels adapted from Csikszentmihalyi (1997). The figure shows the different emotional states that originated from challenge and skill levels. High challenge and high skills indicate a state of flow.*

Furthermore, skills and challenges are the central theoretical criteria of flow because they can originate in different psychological states (Strati et al., 2011; Fullagar et al., 2013; Hamari et al., 2016). For instance, the individual can feel apathy when low challenge and low skills exist, or the person can feel anxious with high challenge and low skills, or relax when the challenge is low, and skills are high (Fig.2.10). Consequently, some people can perceive the challenges as arduous and unpleasant; conversely, others people can feel disengaged in the absence of challenges (Hamari et al., 2016). The optimal state of flow is high challenges and high skills, where the person needs to continuously improve and refine one's skills to achieve growing challenges; and vice-versa, the person increases challenges of the task to enhance one's skills (Strati et al., 2011; Fullagar et al., 2013; Csikszentmihalyi et al., 2014). In the case of arousal state, the individual can move into flow by putting more effort into developing skills. Therefore, in this state, people learn because they leave their comfort zone, trying to achieve the goal (Strati et al., 2011).

However, Cheng et al. (2015) consider that flow is not the appropriate state in video games-based learning because it is an extreme state of concentration and enjoyment. Instead, they propose a suboptimal and non-extreme state called "immersion," which is a precondition to flow. In a real-life situation, a person is highly engaged in playing the video game, but the person needs to leave the game to go to school or catch a bus (Jennett et al., 2008). Thereby, the immersion state maintains engaged experience in the game-play while players are retaining some awareness of their surroundings.

To achieve a complete immersion state, the player goes along three stages (Brown and Cairns, 2004; Bianchi-Berthouze et al., 2007; Cheng et al., 2015). The first stage is engagement which refers to players' preferences; whether the player likes the game, he will continue; otherwise, he will not try to play it. Once the player is engaged, the second stage called engrossment starts. During engrossment, player attention is focused on the game; thus, the perception of the surrounding area and physical needs start to decrease. Moreover, emotions arise with the game; even when players stop playing, they can feel drained. The final stage is total immersion, where players feel involved in the game, and it is all that matters. In this stage, players can feel a virtual presence with affective, cognitive, and emotional components.

In the case of music, to obtain the full immersion (trance) (Becker, 2004, 2022; Maes et al., 2016) the person starts with the first step called absorption, where the musical stimulus occupied the attention of the people without requiring any mental effort. Dissociation is the second step; here, the person employs mentally cutting off from his or her surroundings and external thoughts. In these two steps, the person feels pleasant and effortless. Finally, the third step is trance; this step is an intense experience with the music; the music completely absorbs the attention of the person, focusing only on the music (Becker, 2004). Persons in this third step take voluntary control over the physiology of emotional arousal, originating the suppression of some processes such as reduction of pain or fatigue (Becker, 2022). However, either in total or immersion, learning can be compromised on how cognitive resources are focused and used during learning, also known as cognitive load theory (Chandler and Sweller, 1991). Cognitive load is limited and can be overloaded with the use of complex game dynamics instead of using the cognitive load to process learning content.

Cognitive load theory focuses on the design of learning materials concerning learners' cognitive capacity, and it does not recognize the role of emotional variables (Moreno, 2010) as virtual presence does. During virtual presence, some learning outcomes such as retention and comprehension can be present, but it is not enough for more complex learning outcomes (Schrader and Bastiaens, 2012). In consequence, feeling complete immersion can affect learning, and cognitive overload in the game aspects to get more immersion may inversely result in ignoring educational targets (Cheng et al., 2014).

Furthermore, Baker et al. (2010) complement the emotions involved during learning with emotions such as boredom, confusion, delight, engaged concentration, frustration, and surprise. Although some of these emotions are related to flow condition, they do not consider some aspects of flow, such as clear goals, balance challenges, or immediate feedback. For Baker et al. (2010), these are not only emotions; instead, they consider them as cognitive-affective states because they have significant cognitive and affective components in the context of learning. The affective components are integrated by valence (a pleasure to displeasure) and arousal (activation to deactivation) (Fig. 2.11).

Nevertheless, the effects on learning originated by these states depend on transitions between states and depend on the persistence over time of each state. In the case of systems-based learning, it is important to know about the behavior of the states

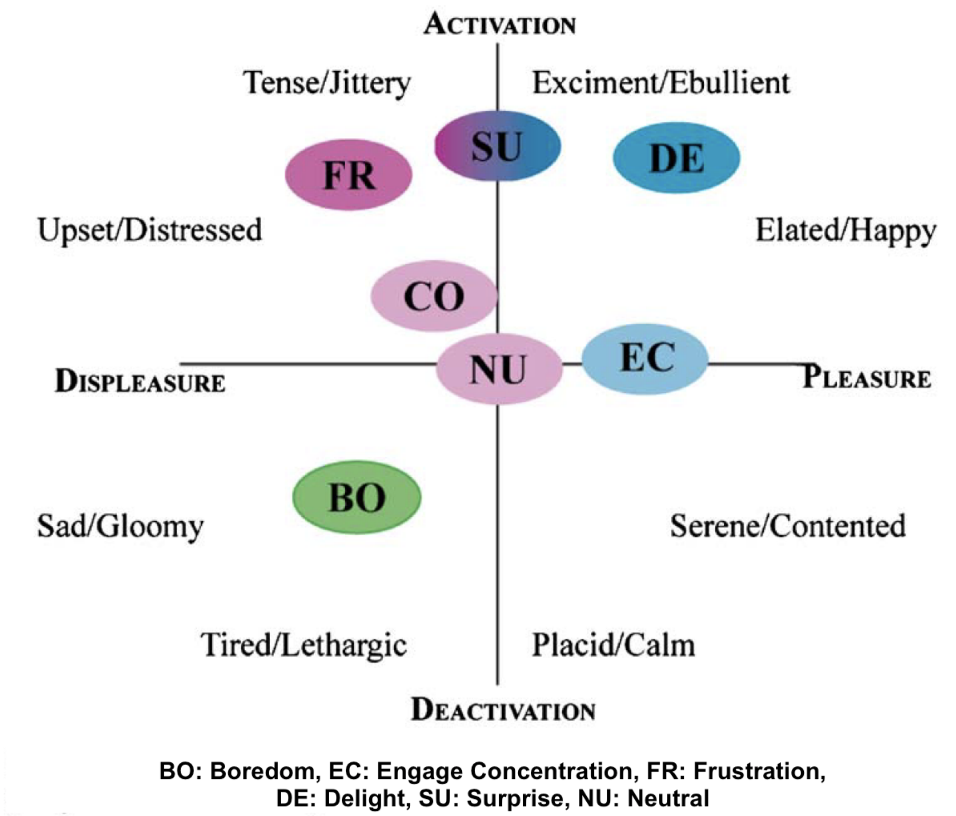


FIGURE 2.11: *Image taken from Baker et al. (2010). Over the Core Affect Framework of Russell (2003), Baker et al. (2010) adapted the learning-centered cognitive-affective states.*

in a learning environment. The states transition proposed by Kort et al. (2001), move from confusion to frustration, and from confusion to boredom. Furthermore, Perkins and Hill (1985) suggested that frustration leads to boredom. In an analysis of some computer-based learning environments, Baker et al. (2010) found that the most thriving states are engaged concentration and confusion, confirming the significant role of confusion in complex learning. In confusion state, persons are in an arousal channel with a cognitive disequilibrium but with more profound thoughts. Additionally, boredom state is non-transitory in most of the cases; once a person is bored, it is difficult to change to another state. One of the reasons is that during boredom state, the person starts to game the system, whereby users systematically guess or abuse hint features to perform well in an interactive learning environment without learning the material. Consequenctly, the person demonstrates poor learning during gaming the system. However, human-computer interaction has has been proven more of the effects of frustration rather than boredom, even when frustration is not entirely correlated with poor learning. Gee (2007) considers that a certain level of frustration can improve the enjoyability of video games.

An important part for learning, because of its direct impact on motivation, are the rewards. The impact of the expected reward depends on the relationship between the reward and the task (reward contingencies) (Deci et al., 1999). McKernan et al. (2015) and Goh et al. (2015), concluded that while a higher number of rewards did not influence the learning in an educational game, players responded more positively to the game when they felt more rewarded. Moreover, rewards also increase the positive affect, enjoyment, and autonomy (Bowey et al., 2015). Johnson et al. (2018) found that rewards impact on presence, immersion and enjoyment, but not competence and that rewards affect the effort, but do not impact on tension.

King et al. (2010) consider rewards and punishment as a reinforcement tool and they are present within the game in different forms such as meta-game rewards or event frequency rewards (for a complete list and examples see Appendix C). Although some elements in video games are associated with positive reinforcement, some elements of failure and punishment are essential in order to establish the contextual worth of in-game rewards, and show the player that making progress is not inevitable but skill-based (King and Delfabbro, 2009). For instance, game developers incorporate failure scenarios such as having to entirely restart a video game when the player's character died or suffered from decreased "health" during an successful attack from an enemy (Kent, 2001). King et al. (2010) suggest that how a player is rewarded for playing a video game is more important than the rewards themselves. Moreover, Hao and Chuen-Tsai (2011) propose four reward system attributes: social value, gameplay affectation, suitability, and time. Social value refers to social interaction for comparison purposes; the second attribute is how rewards affect the gameplay providing new content to proceed in the game. The third attribute is the suitability of rewards for collection and review, with players having motivations that range from building a sense of accomplishment to preserving game memories (Fu, 2016; Formanek, 1991). Finally, the fourth attribute is the time required to earn or receive a reward. Moreover, rewards can change the player's subjectivity, modifying player's emotions through balancing challenge and skill, having clear goals, and immediate feedback.

2.4 Structure of Music Video Games

The video games industry has found an important niche in music. To date, a broad gamut of music video games exists, where music has different roles depending on the specific type of music video game (Pichlmair and Kayali, 2007). Liebe (2013) defines three categories of interaction between player and music: linear, reactive, and proactive. A linear music video game is the one where the music is attached to a specific game-element, and the player cannot influence the video game. For instance, soundtracks of games such as *Heavy Rain* (Quantic Dream, 2010) and *Civilization IV* (Firaxis Games, 2005) are in this category; the music recurs incessantly without any direct influence of players' actions (Fig.2.12). In reactive music video games, music is directly connected to the actions of the player, and specific micro-actions trigger the game-music. Thus, the music depends on the player-actions, for instance, when an avatar changes to fight mode then the music changes. In this category, role-games prevail, such as *Gothic* (Bytes, 2001), and *The Elder Scrolls: Morrowind* (Bethesda Game Studios, 2002). Other reactive music games are *Elektroplankton* (Indieszero, 2005) or *Moondust* (Lanier, 1983) because music is generated through the action of the player (Fig.2.12). Finally, in proactive music video games, players undertake a specific action according to the music. Proactive music is an index for a game belonging to the music-game genre, and rhythm games are representative of this category, such as *DanceDance Revolution* (Konami, 1998) or *Patapon* (Pyramid and SCE Japan Studio, 2007) (Fig.2.12).

FIGURE 2.12: *The figure displays three examples of video games according to the categories proposed by Liebe (2013). The image shows as an example of linear games, the soundtrack (CD), and the cover of the game Civilization IV (Firaxis Games, 2005) as an example of linear games. In the category of proactive games, the image displays a screenshot of the game Dance Dance Revolution (Parodius8662, 2009). The example of reactive games is represented with the screenshot of the game Gothic (RedFox, 2016).*

A critical point between video games and music is the concept of interactivity, defined by Seifert (2008) as ”to act upon.” Two different things interact if, and only if, each acts upon the other. The actions of all agents, be they human, system, or machine, have to influence one another to establish a process of interaction.

Pichlmair and Kayali (2007) analyzed several music games qualitatively, regarding how games structure interactivity, and they classified them according to common features (See Table in Appendix D.1):

1. **Active score music** refers to a concept for a musical score that can be adapted to each performance. It creates a dynamic soundtrack for the game.

2. **Rhythm-action** is based on repetitions and variations, offers little freedom of expression, and contains restrictive rules on how the player has to react.

3. **Quantization** considers that the action of the players should be aligned to the beat of the game. The challenge is not to hit the beat but to create a steady flow of sound.

4. **Synaesthesia** is defined as an involuntary neurological state in which different sensations are coupled (Cytowic, 1989); for instance, visual metaphors for audio are cued. The gameplay seeks to attract both the visual and sonic sensors at the same time.

5. **Play as performance** is a game mechanic in which the physical performance of the players is vital. It allows for a wide range of expression and bodily engagement.

6. **Free-form play** is characteristic of digital toys rather than games. Since goals are not present in a game, such as hitting something, it turns a game into a toy or instrument. Some games, in some modes, allow the exploration of sounds and music without a specific goal.

7. **Sound agents** indicate the presence of specific game elements that enable interaction with sound but have a behavioral pattern on their own. Music is generated while exploring the function and characteristics of these agents.

Based on the analysis of Pichlmair and Kayali (2007), Liebe (2013) establishes two categories of music games. One category is computer performance, where sounds support the involvement of players in the game and also give players a meaning of the game. In this category, it is important to generate an atmosphere created by the use of typical instruments or sounds of a specific region and extensive use of musical clichés, e.g., *Brütal Legend* (Double Fine Productions, 2009). The involvement and meaning of the game depend on acoustic identities that refer to sounds characteristic of the game, that identify each game, such as in *Star Wars*[3] or *Super Mario* (Nintendo EAD and Nin-

[3] Lucas, George, dir. Star Wars Episode IV: A New Hope. Twentieth Century Fox, 1977. Film.

tendo EPD, 1985). Moreover, sounds in computer performance rely upon ergo-audition, which are self-generated sounds such as kicking a tin can. These sounds involve players the most as it relies entirely on their actions. Thus, players feel actively involved in the sound environment of the game. Ergo-audition is a representative element of the reactive sound structure mentioned above. The other category of Liebe (2013) is player performance, where the player performs more active physically than virtually. For instance, *DanceDanceRevolution* (Konami, 1998) uses a mat with sensors, where players must jump to match the visual cues of the game rather than only press buttons or move a joystick. In these games, the mechanical aspect of the game is fun and straightforward, whether or not the physical actions align with the virtual aesthetic.

Interaction between player and game is established through a hardware interface, which is the link between images, sounds, and music, and the player. To date, new interfaces for music games consider gestures as an important part of a computer game interface design. Microphones exist in today's market which enable players to sing while playing games such as *Singstar* (London Studio, 2004) or *Rock Band* (Harmonix and Pi Studios, 2007)) or the *Microsoft Kinect* (Microsoft, 2010). In addition to microphones, cameras enable the tracking of player' movements in games such as *Dance Central* (Harmonix, 2010)) as well as plastic devices (instruments) to imitate a real musical intruments such as *Guitar Hero* (Harmonix et al., 2005) or *DJ Hero* (FreeStyleGames and Exient Entertainment, 2009). Moreover, the mobile versions of these interfaces also consider gestures by standard controllers or touch-screens, although they are far less performative.

Finally, Liebe (2013) proposes two categories to classify interfaces in music video games according to object interaction: objective and symbolic. In the objective category, the gadget that the player uses to interact represents what the player can do with the object. On one hand, in the symbolic category, the gadget does not represent what the players are doing. For instance, in *Guitar Hero* (Harmonix et al., 2005), what players do with the controller is the same as presented on the screen (objective); whereas in *Super Mario* (Nintendo EAD and Nintendo EPD, 1985), the player presses a button to jump, the player is not jumping, rather a button symbolizes the jump. Similar to Liebe (2013), de Castell et al. (2014) refer to two classes of interfaces: imitation and simulation. The simulation uses the phrase ”How is” real (Jenson and De Castell, 2009), e.g., *Super Mario* (Nintendo EAD and Nintendo EPD, 1985), where the electronic device or control has buttons to press to make Mario jump, but you do not need to jump. Furthermore, imitation uses the phrase ”as same that” real (Jenson and De Castell, 2009), e.g., the use of the *Kinect* (Microsoft, 2010), where you need to jump physically to make an avatar in the game jump. Jenson and De Castell (2009) and de Castell et al. (2014) researched the use of these interfaces for educational purposes. They found that people who play and do not have musical education can learn to play guitar easier than people who have never played. These controllers train the muscles and some abilities of gamers (Jenson and De Castell, 2009) and origin muscle memory (Jäncke, 2006). For instance, the muscle of a drummer is exercised during a *Guitar Hero* (Harmonix et al., 2005) session, and it causes muscle memory so, when the person starts to learn guitar, it is easier because he or she has this memory in his or her body.

According to Liebe (2013), *PaRappa the Rapper* (NanaOn-Sha and Japan Studio, 1996) belongs to the category of computer performance. Interaction with the game is through a traditional controller (Fig. 2.13-A); thus, this game also belongs to category symbolic interface because buttons represent the actions, the player has to press the buttons to perform the actions of the game, e.g., R1 to jump. Moreover, the game is located as a proactive game because the player undertakes a specific action when sound is played. On the other hand, *Donkey Konga* (Namco, 2003) is a proactive game because the player hits the ”conga” interface at a specific time (Fig. 2.13-B). However, different than *PaRappa the Rapper*, this game has a specialized electronic interface for the game (a ”conga” interface); thus, it is classified as an objective interface because the player gives a real hit to the *Konga*. Thereby, the game belongs to the category of player performance. Whereas, in a different category, *Batman Arkham* (Rocksteady Studios et al., 2009) is a computer game, which plays by the use of a traditional controller where actions are symbolic (Fig. 2.13-C).

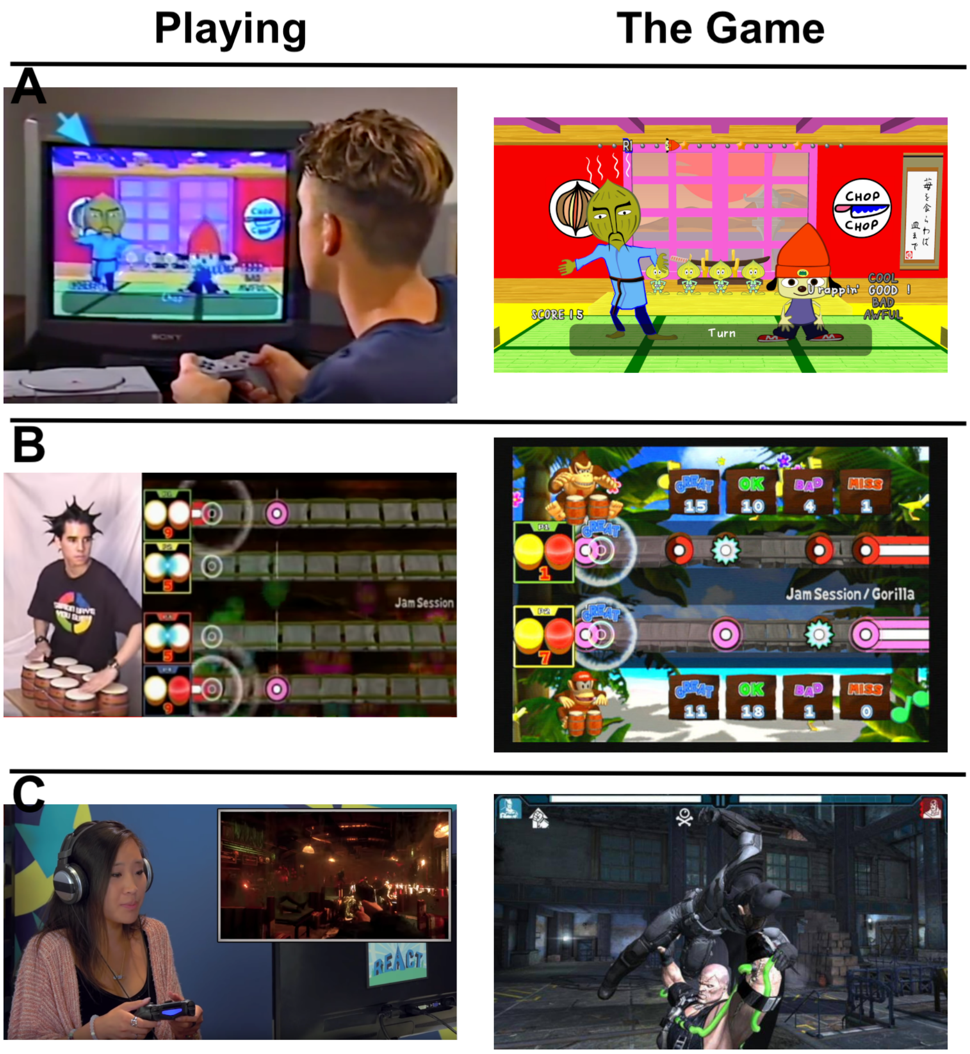


FIGURE 2.13: *Left column: a player using the controllers of the game during the gameplay. Right column: a screenshot of the game.* ***A)*** *A screenshot Pips (2017) of a player using a tradition controller during the gameplay of PaRappa the Rapper.* ***B)*** *Screenshot CrackedRabbitGaming (2006) of a player hitting the Kongas of Donkey Konga game.* ***C)*** *A screenshot React (2015) of a player playing the game Batman Arkham.*

This game is classified as a linear game because each sound is attached to a specific game-element, and the player cannot influence the sounds.

2.5 Learning Music with Video Games

Children have their first contact with music through toys, video games, and other devices than with an authentic musical instrument. According to the National Foundation for Youth Music UK (Missingham, 2007; Gower and McDowall, 2012), youths believe that music video games can provide access to music education for more people, independently of their social stratum. Furthermore, music video games can motivate people as well as have an important role in education (Gower and McDowall, 2012). Including interactive music video games in the curricula can result in an exciting and meaningful experience for young people and teachers. Games motivate players and also support learning or training new skills, such as certain body movements. Learning through a video game is not directly associated with the content itself, in the sense of information rooted in intellectual domains or academic disciplines. Technologies offer new ways of communication, where reading and writing text is not the only communication system, but also, images, graphs, or diagrams have meaning (Gee, 2003). Different school disciplines take advantage of the use of literacy to provide meaning, representations, and sensations to students (Eisner, 1991). For instance, literacy in traditional music learning focuses on reading and writing standard music notations; however, the use of new literacy (e.g.,visual stimulation through video games) is a key point to realizing the potential of video games in music lessons as a new way of reading and writing (Gee, 2007; Eisner, 1991). Kylie et al. (2011) showed that players who are familiar with a game-specific notion of fundamental music concepts (e.g., meter or rhythm) internalize them and have better scores in traditional music assessments. Therefore, music educators might consider the interaction between students and this media to create meaning and develop music skills (Tobias, 2012).

Learning through music video games also requires cognitive abilities such as memory, reaction time, attention, and selective attention. For instance, auditory, visual or tactile coordination is important in order to be accurate in rhythm video games. In their study about the use of real-time visual feedback during singing lessons, Welch et al.

(1989) found that participants improved their singing using visual feedback; moreover, this improvement rose when a combination of visual feedback and verbal feedback (teacher presence) existed. Nevertheless, the design of games for learning and games for entertainment vary according to the target. In games for learning, the design is based, mainly, on the drill-and-practice modality. These games provide repetitions; therefore, the players practice the task until they acquire knowledge and music skills (Chung and Wu, 2017). Recent work proposes that music video games, such as *Guitar Hero* (Harmonix et al., 2005), can help students improve their musical skills and support the embodied competence of players. For instance, people who play music video games and do not have a musical education can learn to play guitar more easily than people who have never played such kind of video games (Jenson and De Castell, 2009; de Castell et al., 2014). These music skills are acquired through the repeated use of computer interfaces, which have an impact on players' muscles and create muscle memory (Jäncke, 2006; Jenson and De Castell, 2009; de Castell et al., 2014). Learning through video games is associated with activation of game elements and a challenging situation faced by the player (Gee, 2007; Jenson and De Castell, 2009).

Understanding how music is represented in a multimodal context of video games allows new ways of perceiving and interacting with music. Thus, it is important to start correlating the body, sound, and video game to obtain a music experience during the gameplay and after through visual, aural, or other stimuli. Tobias (2012) considers five modes that video games and gameplay have for music teaching and learning:

- The first mode consists of allowing video games to coexist with other tools in classrooms such as instruments, textbooks, space, or other resources. The aim is to create a connection between the virtual and real worlds.

- The second mode needs to create affinity groups where students explore musical and gaming cultures (Gee, 2003; Squire, 2008).

- In the third mode, music educators might build on the affordances of video games, treating their constraints as learning opportunities.

- The fourth mode considers multimodality and interactivity as an effective strategy to create music meaning and schema in terms of space, structure, action, and other concepts.
- In the fifth mode, music educators might act as facilitators without interfering in their work and construction of understanding (Hargreaves et al., 2003; Gee, 2007).

2.6 Chapter Summary

Body movement and perception of some stimuli are elementary for music learning. The first section of this chapter described how the embodied actions are associated with the cognitive process in two aspects: ground cognition and extended cognition. The ground cognition is based on sensorimotor networks, whereas extended cognition considers the environment and the external factors. These two aspects drive into a more action-oriented view in cognitive science. Afterwards, the section was linked with a description of the mirror neuron system which plays a central role in the cognitive process. This section focused on the goals and intentions of the players' actions and how observation affects movement execution. The following section considered the concepts previously pointed out in the music field. Two main components of music (rhythm, and pitch) were analyzed from a cognitive base. At this point, the chapter addressed the use of technology, video games in particular, for music learning. Therefore, the following sections will describe how cognitive processes affect player performance in music video games. The concepts pointed out in this chapter are the bases of the creation of a framework to design music video games, considering the cognitive processes of the human body.

Chapter No. 3

Rhythm as the Main Element of Music

This chapter focusses on a prototype based on rhythm structures. The first part of Chapter 2 describes the prototype design, such as the rhythm sequences used for a game and its levels. Moreover, this section explains the use of some game-elements. Afterward, the chapter presents two case studies that use the rhythm prototype presented in the first part. The first case study analyses players' behavior in a synchronization task using either auditory or visual stimuli to reproduce the rhythm sequences. The second case study employs tactile stimuli to reproduce the rhythm sequences compared to auditory stimuli. The performance of the players was analyzed during a synchronization task.

3.1 Introduction

To date, in rhythm video games, it is not clear which stimuli affect the players' performance or which stimuli attract of more the players' attention. Furthermore, game-elements can act as distractors but, at the same time, as a supportive element. These constraints will depend on human perception and sensorimotor systems. However, there is no reference for designers regarding which game-elements can be added to the game or which stimuli are more suitable for specific target groups. For music learning, these considerations are essential in order to enhance rhythm skills in the players. Designers need to balance game-elements and game-mechanics with humans' perception and the sensorimotor system. Therefore, some questions arise: 1) in a video game environment, which stimuli does the synchronization task enhance? 2) Do the auditory stimuli improve the player experience and reduce the workload in players? 3) How much do visual elements, such as platforms, affect the player performance following certain stimuli? To find an answer to these questions, a rhythm game was developed in two versions. The first version presents rhythm sequences using either auditory or visual stimuli. In contrast, in the second version, the rhythm sequences present either auditory or tactile stimuli. Both versions consider two levels of difficulty. Furthermore, each version of the prototype presents a case study that evaluates the players' behavior in a synchronization task. Both case studies validate or reject the hypotheses mentioned below in each section.

3.2 The Design of the Rhythm Prototype

The design is a 2D side-scroller jump-and-run game based on rhythm patterns (sequence) with two levels of difficulty, where the goal of the players is to follow the patterns. Thus, all interaction in the game is driven by the underlying music in the game. The core mechanic consists of an avatar jumping from platform to platform in synchrony with the rhythm patterns (sequences). There are five different sequences with a length of not longer than eight beats considering quarters, eights and silences (Fig.3.1). The tempo of 90 beats per minute (bpm) is used to synchronize the side scroll.

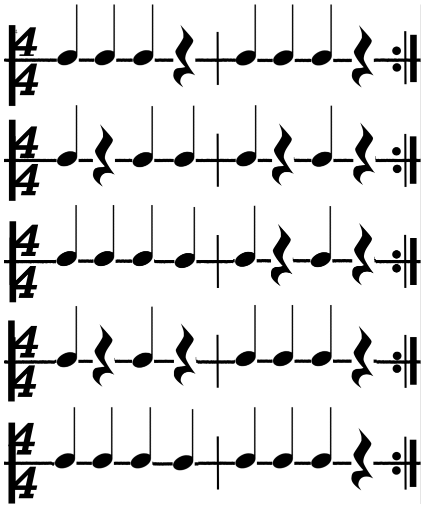
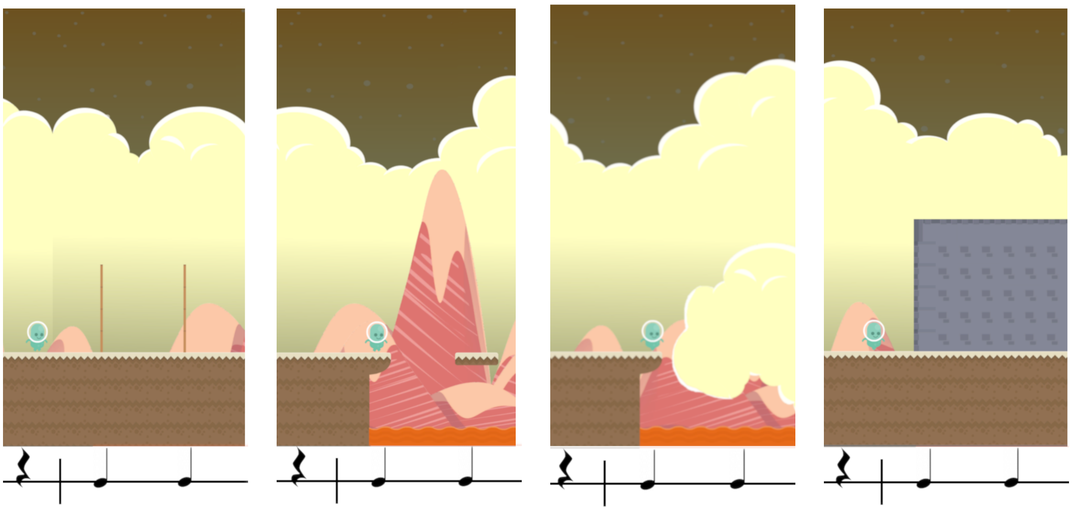


FIGURE 3.1: *Sequences of rhythm patterns used in the game*

The patterns are visually represented by bamboo sticks, platforms, and gaps in the game. Therefore, Level 1 presents each sequence two consecutive times. The first time is represented with bamboo sticks (Fig. 3.2a). When it is repeated, the patterns appear in the form of platforms and gaps (Fig. 3.2b). In Level 2, each sequence appears four times: the first time is the bamboo sticks area (Fig. 3.2a); the second time is the platforms and gaps part (Fig. 3.2b); the third time is the bamboo area again, (Fig. 3.2a); and in the fourth time, clouds or walls hide the platforms and the gaps (Fig. 3.2c - 3.2d).

Regarding the interaction of the player with the video game, the player controls the avatar by pressing the space bar. To obtain a successful jump, the player must press the space bar in synchrony with the rhythm. Double jumps are not supported. If the player does not perform a jump at the exact timing of the rhythm pattern, the character cannot reach the next platform and falls into the gap, and the player loses a life. For each level, the player gets three lives to reach the end of the level. As a consequence of losing a life, the player can repeat the exercise departing from the sequence where the player has lost the life. A score point system rewards a successfully performed exercise.



(A) *Bamboo section* (B) *Platforms* (C) *Platforms hidden with clouds* (D) *Platforms hidden with a house*

FIGURE 3.2: *The game Jump'n'Rhythm (Alexandrovsky et al., 2016).*

The aim of the design is that players listen to the rhythm in the bamboo area and memorize or internalize it without performing any action. Then, they should reproduce the same sequences, but now by pressing the space bar in synchrony with the rhythm sequence. The goal of Level 1 is to also internalize the rhythm sequences, because the sequences in Level 2 will be presented in the same order. In Level 2, however, difficulty increases due to hidden platforms and gaps. This design of Level 2 forces the player to perform according to the stimuli and not focus only on the platforms.

The game presented in this section was used to carry out two case studies which are described below. The first case study evaluates the use of auditory and visual stimuli in the rhythm prototype, whereas the second one employs auditory and tactile conditions in the game.

3.3 Using Auditory and Visual Stimuli in a Rhythm Game

Synchronization between a player's body movements and the video game environment is fundamental since video games require accuracy in the players' reaction time to events or patterns. In this interaction, a player receives information about the video game through the body to make an internal judgment and decide how the human body will react. Most video games transmit this information through stimuli (e.g., auditory, visual). Recent studies (Repp and Penel, 2004; Kato and Konishi, 2006) have found that auditory stimuli dominate over visual stimuli in temporal perception. For instance, synchronizing the index finger with auditory stimuli is more precise than with visual stimuli (Repp and Penel, 2004; Patel et al., 2005; Manning and Schutz, 2013; Hove et al., 2013b). However, when visual stimuli become spatio-temporal, the reaction time enhances (Hove et al., 2013a,b). An auditory dominance is also observed when the synchronization targets are auditory sequences. However, a second sequence (auditory or visual) acts as a distractor (Bootsma and C. W. Van Wieringen, 1990; Tresilian, 1994; Miyake et al., 2004; Hove et al., 2013b). For instance, on one hand, whether the players' attention is more focused on the visual interface than tapping the sequence, the synchronization task becomes difficult. On the other hand, whether the players' attention is only focused on their tapping task, making temporal decisions can affect the anticipatory timing control. Consequently, the attention to the video game's dynamic could be a critical resource for the performance of the player.

Therefore, in the present study, the players' behavior in a synchronization task was evaluated using auditory or visual rhythm sequences admitted by the prototype *Jump'n'Rhythm*. For this study, three hypotheses were stated:

- H1: Players have a better reaction time using auditory stimuli, even in the absence of supportive visual game-elements.
- H2: Musicians have a better reaction time than non-musicians using auditory or visual stimuli.
- H3: The workload of participants using visual stimuli is higher than participants using auditory stimuli.

3.3.1 Case Study

In the present study, the reaction times of the players to sequences of auditory or visual stimuli provided within a video game were examined to analyze which of the stimuli enhance accuracy in their performance. The experiment consisted of two groups of participants who performed a video game. One group primarily followed auditory stimuli, whereas the second group followed visual stimuli, in two levels of difficulty. Better accuracy was expected from the players with auditory stimuli. By comparing these groups, the study considers two main questions: 1) which stimuli enhance the accuracy of players' reactions time in a rhythm video game, for future design consideration? And 2) how does visual design change the performance and accuracy of reaction time of players?

Participants. An invitation was made for participants living in Bremen (Germany) area, in particular for those with a music background. Additionally, employees of a magazine who specialized in music, located in Cologne (Germany), were invited to participate. Most of the employees were professional musicians. In total, the case study included 24 participants (6 females and 18 males) ranging from 17[1] to 41 years old. All of them declared themselves as right-handed persons. Twelve participants were professional musicians (10 males and 2 females). They reported no sensory defects, no physical disability, nor any mental disorder. The participants were split into two groups, one group using auditory stimuli and the other one visual stimuli. The musicians with a music background were divided into two groups of equal size.

Apparatus. The video game *Jumpn'Rhythm* used in this experiment was developed using the Unity[2] engine. The participants played the game running in Unity on a Mac Book Pro with a 2.4GHz Intel Core i5 processor with an NVIDIA graphics board. The screen employed to visualize the video game and to show the visual stimulus was a 17 inch Hanns-G monitor. The screen resolution was 1440×900px, and the game had a

[1]With a parental approval

[2]Unity Technologies. (2005). *Unity*. [Game Engine]. Available at `https://unity.com/`

resolution of 1368×770px. The participants used AKG k271 Studio headphones either to listen to the auditory stimuli or to cover possible external noise in the visual condition. The tapping task was performed on an external mechanical keyboard connected via USB to the computer for playing the video game (Fig.3.3).

FIGURE 3.3: *A participant plays the game in the auditory condition. The monitor shows the video game. The player uses the keyboard to perform and the headphones to listen to the auditory stimuli. The microphone is under the keyboard on the left side.*

The Stimuli. The prototype has two versions, auditory and visual. In both versions, the tempo of the sequences of rhythm patterns was 90bpm (one beat every 667ms). Therefore, the synchronization task was comfortable, mainly using visual stimuli where short intervals might be challenging to synchronize (Repp and Penel, 2004). Metrical sequences in include intervals of only 667ms, considering quarters and silences (Fig.3.1) in a music context.

In the auditory version, the stimulus for rhythm patterns is an audible tone of 1000Hz emitted through the video game. Moreover, the audible tone uses a percussive envelope consisting of an attack time = 0.005s and a release time = 0.110s (Fig.3.4).

A blinking green circle shapes the visual stimulus with a diameter of 15cm. The green circle blinks in the center of the display each time a rhythmic pattern occurs (Fig.3.4).

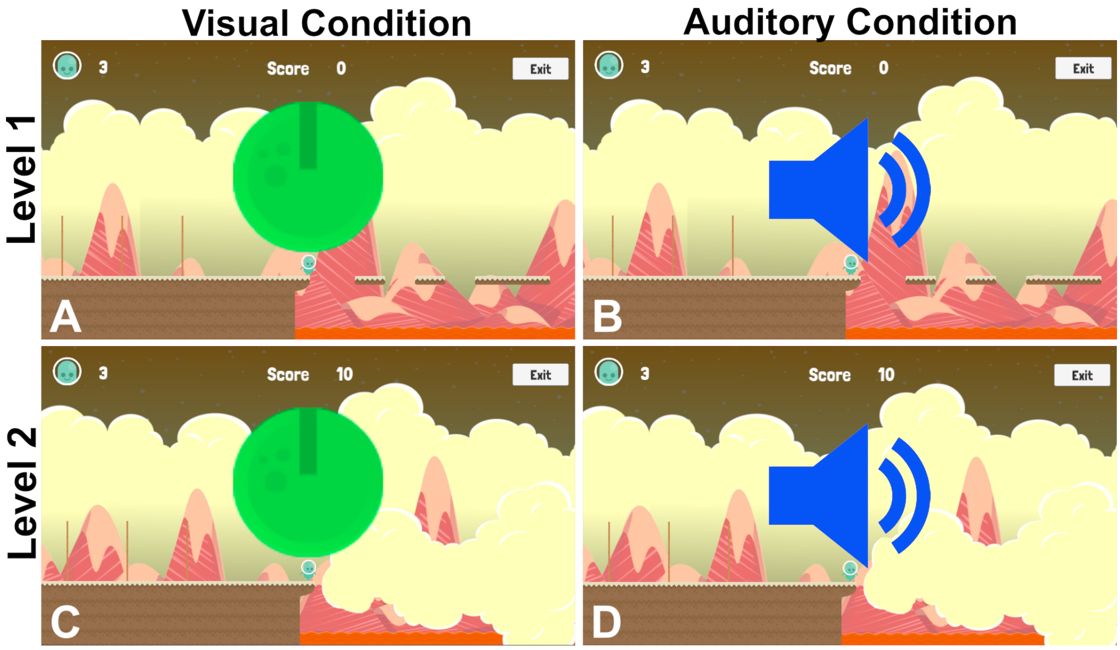


FIGURE 3.4: ***A)*** *Visual condition of the video game scene with the elements and the visual stimuli, at Level 1.* ***B)*** *Auditory condition of the video game scene with the platforms and the auditory stimuli, at Level 1.* ***C)*** *Visual condition of the video game scene with hidden platforms and with the visual stimuli, at Level 2.* ***D)*** *Auditory condition of the video game scene with hidden platforms and with the auditory stimuli, at Level 2. Note that the speaker symbol represents the beep of the auditory stimulus and was not visible.*

Procedure. Participation consisted of only one session in a quiet room with good lighting, where each participant remained seated in front of a monitor and a computer keyboard. The session started with a demographic questionnaire about the players'expertise in gaming and music. Afterwards, participants played the video game with a single task of synchronizing their taps with the (prescribed) beat of rhythmic sequences, auditory or visual. The participants were instructed to tap only once per pattern using the index finger of their preferred hand on the mechanical computer keyboard. Players played Level 1 and Level 2 of the game. The design of the experiment was between-subjects, which means one group performed one and only one condition. Therefore, participants under auditory stimuli followed auditory sequences, whereas participants using visual stimuli followed visual sequences.

In order to evaluate the perceived levels of mental, physical, and time demands associated with the stimuli and Level the participants completed a questionnaire of unweighted NASA task load index (TLX) (Hart and Staveland, 1988). Afterward, following the session, participants completed post questionnaire about their opinions of the game.

Data Processing. Data were obtained following a similar approach to data acquisition than Tierney and Kraus (2015) and Tierney et al. (2017). Their research focused mainly on the relation between rhythm and language skills; for instance, their investigation regarding the difference between beat tapping and rhythm memory/sequences, where they found that these are dissociable rhythmic aptitudes (Tierney and Kraus, 2015).

Therefore, in this study, participants played the video game with the single task of synchronizing their taps with rhythmic sequences, either auditory or visual. For listening to the auditory sequences, participants used headphones; while visual sequences were presented through the video game. The participants were instructed to tap only once for each pattern (auditory or visual) using the index finger of their preferred hand on the mechanical computer keyboard. A passive microphone located close to the keyboard received the sound emitted by the keyboard. The amplified signal obtained from the microphone and the signal from the video game (running in Unity) were routed via Loopback[3] and recorded in a computer using Audacity[4][5]. Therefore, audio from the microphone was sent to the left channel, while the signal from the video game (rhythm

[3]Rouge Amoeba. (2016). *Loopback*. [Computer Software]. Mac Os. USA: Rouge Amoeba

[4]The Audacity Team. (2000). *Audacity*. [Computer Software]. Mac Os, Windows; Linux and Unix. Available at `https://www.audacityteam.org/`

[5]Audacity was calibrated to avoid possible latencies as a result of wiring of microphone and pre-amplifier. In the calibration test, the final obtained latency was of 0.005s.

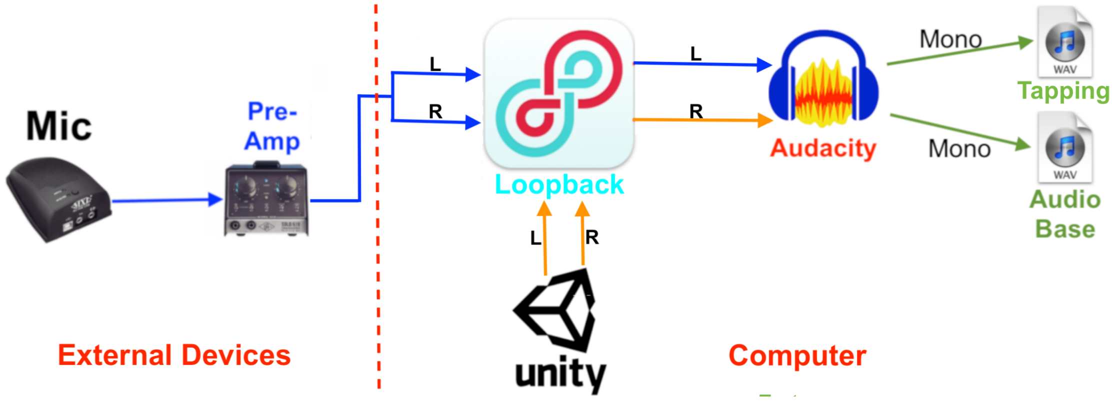


FIGURE 3.5: *The diagram shows the connections and direction of the signals. The signal from the microphone is transmitted via USB to the laptop computer considering two audio channels (left and right) by default in the input signals. The software Loopback routes the audio from the microphone and Unity, and sends the audio to Audacity where the outcome is two wav files.*

sequences signal[6]) was sent to the right channel (Fig.3.5). The sample rate of the audio recording was 48kHz.

Both audio signals were edited in Audacity; this edition consisted of normalize amplitudes, noise reduction, and removal of bamboo parts to keep only the platform areas[7]. All the modifications were done in parallel with both signals to avoid shifting with respect to the time domain. In the end, each signal was saved separately in a wav file. The files were exported to Matlab[8] to continue with the analysis (Fig.3.6).

The first part of the analysis was to normalize the length of the signals by padding them with zeros in time-domain. The envelopes of both signals were computed using the Hilbert transform, where the signal peaks were connected to obtain their outline. From the envelopes, the analysis considered only peak values higher than 0.7, and maximum peaks of signals were obtained (Fig.3.7).

[6]The rhythm patterns of the game emit the ”rhythm sequences signal, which is recorded in Audacity either for the auditory stimuli or for the visual stimuli. For the auditory condition, the participants can listen to this ”rhythm sequences signal”. For the visual condition; however, ”rhythm sequences signal” was muted, but Audacity still received the signal of ”rhythm sequences signal”.

[7]Participants varied in their performance during the bamboo area, some of them jumped, and others only listen to the rhythm without jumping. Therefore, only the part of the platforms was considered because in this part, all the participants were forced to jump.

[8]MATLAB 2018a, The MathWorks, Inc., Natick, Massachusetts, United States.

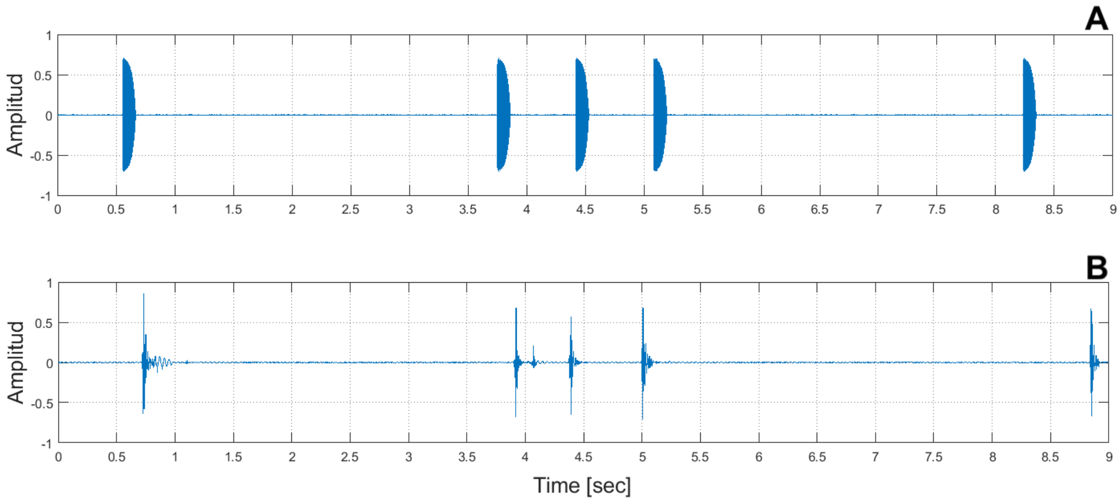


FIGURE 3.6: *The figure shows a segment of the signals with rhythm patterns; the range of the segment is the same in time-domain at both signals. In **A)**, the signal belongs to the audio emitted in Unity, which corresponds to the rhythm sequences signal, either auditory or visual stimuli. Whereas, in **B)**, the signal belongs to the audio from microphone (i.e., audio of the tapping task of the participants).*

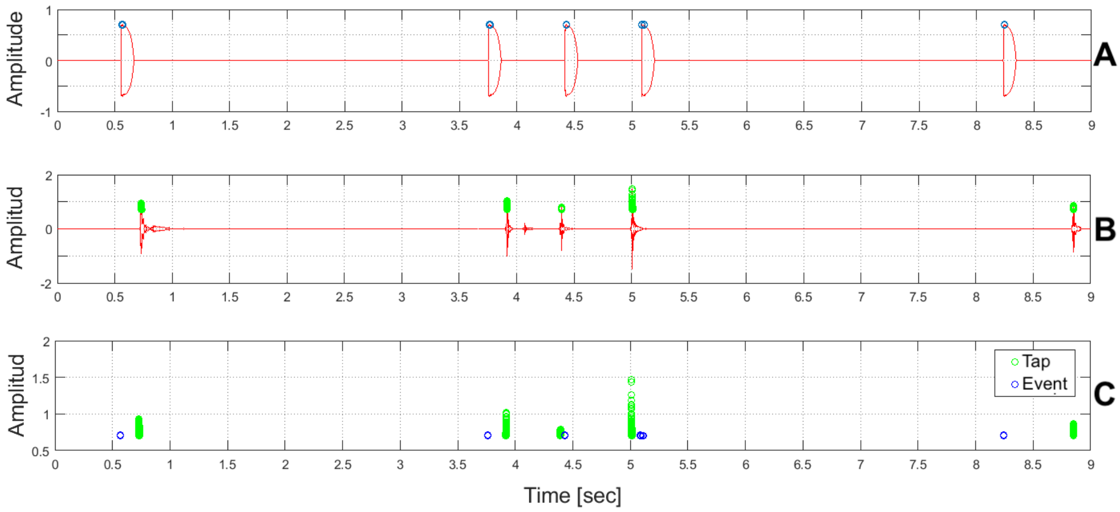


FIGURE 3.7: ***A)** Graph of envelopes for the audio signal. Peaks higher than 0.7 are displayed in blue. **B)** Graph of envelope for the tapping signal. Peaks higher than 0.7 are shown in green. **C)** This graph displays the peaks higher than 0.7 for both signals, in blue for audio and green for tapping. Note that a time difference between peaks of the signals can be observed.*

Finally, a comparison was done between the maximum peaks of the recorded signals (tap) and the maximum peak of rhythm sequences signal. Thus, time differences between the rhythm sequence signal and tap of the player were calculated (Fig.3.8). For instance, from Fig.3.7-C, an event (stimuli) occurred in the interval 0.5s-1s; thus, the blue circle (tapping) is 0.57s and the green circle (rhythm sequences signal) is 0.74s. Therefore, there is a time difference of -0.17s for that event (First blue point of Fig.3.8). This result means that the player did not have perfect synchronization (0s); instead, the player had a rush of 0.17s between his or her reaction for this pattern and the perfect accuracy (0s). It depends on how far the player went during gameplay is the number of pattern that were registered per player, which means the amount of time differences in seconds per player, e.g., Fig.3.7-C had five patterns, which means five time differences in Fig.3.8.

Consequently, the analysis of the study showed there were set or vector time differences for each player. Therefore, twenty-four vectors, twelve for the auditory condition and twelve for the visual condition were obtained.

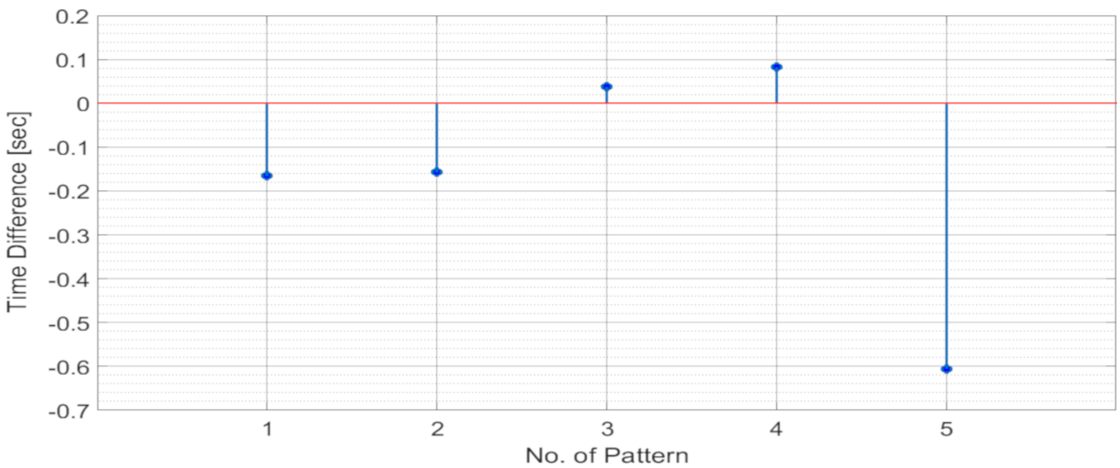


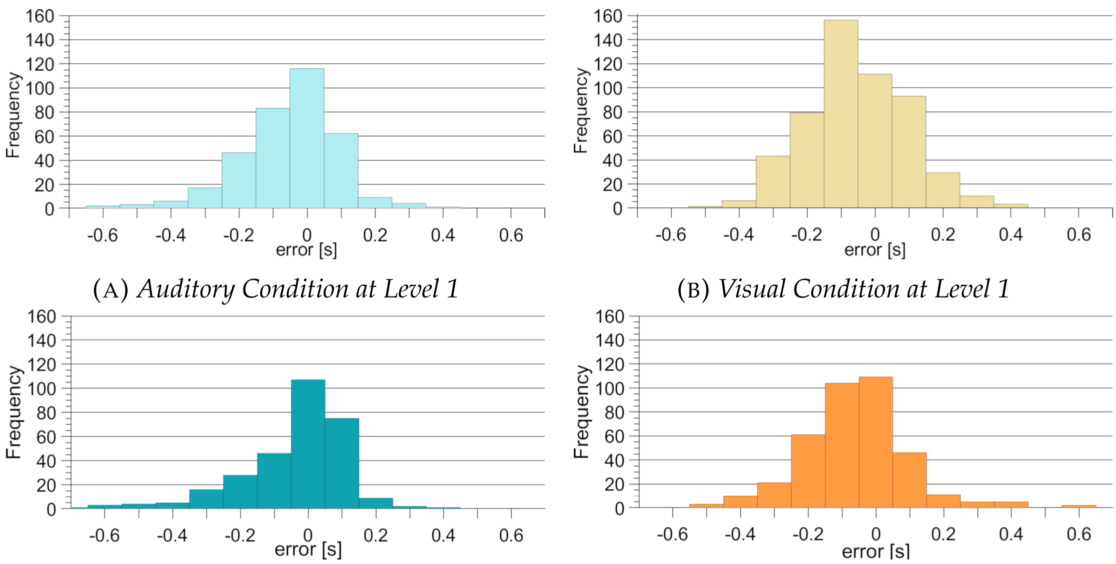
FIGURE 3.8: *The blue points represent time differences between the rhythm sequences signal (stimuli auditory/visual) and tapping signals (player tap). A blue point located in zero (red line) indicates that the player taps in perfect synchrony with the stimuli.*

3.3.2 Results

Results of the study are split into two sections; the first section (3.3.3) analyzes the data considering only two groups: auditory condition and visual condition, without taking into consideration the music background of participants. In the second section (3.3.4), the data were divided into four groups: musician-auditory, non-musician-auditory, musician-visual, non-musician-visual. In both sections, data were analyzed by level and condition.

3.3.3 Results Discarding the Music Background of Participants

Histograms represent graphical data to visually identify where variation exists. They display the frequencies of a data set, which are useful when wide variances exist among observed frequencies for a particular data set. The vertical lines represent how many times tapping falls into a range of time. In the data set of the study, histograms display



(A) *Auditory Condition at Level 1*

(B) *Visual Condition at Level 1*

(C) *Auditory Condition at Level 2*

(D) *Visual Condition at Level 2*

FIGURE 3.9: *The histograms show the distribution of the data for all the participants in each condition. The positive values indicate that the player pressed the space bar of the computer keyboard before the stimuli appeared. Negative values indicate that the player pressed the space bar after the stimuli, indicating a reaction.*

how many taps players performed close to 0s, which means perfect synchrony between the stimuli and the tap. Moreover, information about the tendency of participants to anticipate (negative values) or to delay (positive values) the taps to the stimuli can be observed.

In Level 1, histograms show that most of the taps of players using auditory condition occurred around 0s (Fig.3.9a). In contrast, with the visual condition, most of the taps appeared around -0.1s (Fig.3.9b). Meanwhile, for Level 2, histograms show that taps of players with the auditory condition were in a tighter range of -0.1s to 0.1s (Fig.3.9c in comparison with a broader range (-0.2s to 0.2s) for the visual condition (Fig.3.9d).

However, delays or anticipations of the players' taps were not the value of interest in this study. Instead, the value of interest was accuracy in terms of the magnitude of time; therefore, data were considered in absolute values (of the time differences), also referred to as absolute accuracy. To compare the distribution of absolute accuracy between stimuli and levels, a summary of the distribution is more suitable to avoid distraction with other details such as the overall shape. Thus, a boxplot helps explain where the distributions lie concerning one another by plotting the quantiles side by side.

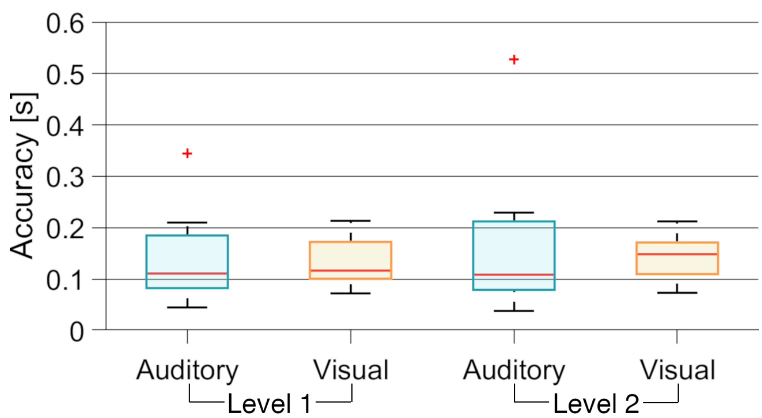


FIGURE 3.10: *The boxplot shows the dispersion of the absolute value of data from the participants, based on the minimum and maximum values, median, first and third quartile, and possible outliers.*

The boxplot for Level 1 shows absolute data distribution with a median of 0.110s for the auditory condition, whereas for the visual condition the median was 0.1157s (Fig.3.10). In addition, an outlier of 0.3442s was found with the auditory condition. For Level 2, the boxplot shows a median of 0.108s for the auditory condition and a median of 0.148s for the visual condition (Fig.3.10). Moreover, an outlier of 0.5281s was found. For the subsequent analysis, these two outliers (i.e., data that are out of tendency) were removed.

In order to identify in which condition participants performed closer to a perfect synchronization, the Root Mean Square Error (RMSE) of the absolute value (of time differences) was calculated per level and condition. The RMSE is the standard deviation of the residuals[9] (prediction errors) (Glen, 2013). It is a measure of how concentrated the data is around the line of best fit. RMSE is a good measure of accuracy to compare residuals of different models for a particular variable but not between variables. RMSE is commonly used in forecasting, and regression analysis to verify experimental results.

Thus, RMSE was calculated through:

$$\mathbf{RMSE} = \sqrt{\frac{1}{N}\frac{\sum_{i=1}^{N}(P_i - O_i)^2}{\delta d_i^2}} \tag{3.1}$$

where N represents the total number of taps reached by the player, P_i represents predicted value when the participant taps, O_i represents the time when an event of the rhythm sequences signal occurs, and δd_i is the point of data-error[10]. A RMSE = 1 corresponds to an optimal fit within data-error. A value of 0.001s was proposed as data-error, which corresponds to possible latencies.

The results of the calculated RMSE per level and condition are shown in Table 3.1, where the RMSE, in Level 1, for auditory condition was of 14.2635, unlike the visual condition that was 15.5350. Meanwhile, for Level 2, the auditory condition had an RMS of 15.4310, in contrast to the RMS of 16.5543 for visual condition. Subsequently, absolute mean accuracy and the standard deviation were computed for both levels and

[9]Residuals are a measure of how far from the regression line data points are

[10]Data-errors would correspond to a persistent bias coming from mis-calibration

conditions. Therefore, for Level 1, auditory condition displays a mean of 0.1208s and a standard deviation of 0.0568s, while the visual condition shows a mean of 0.1325s and a standard deviation of 0.0428s (Table 3.1). Level 2 presented a mean for the auditory condition of 0.1219s and a standard deviation of 0.0657s. Whereas the mean for the visual condition was 0.1413s and a standard deviation of 0.0451s (Table 3.1).

TABLE 3.1: *Results of Mean, Standard Deviation (SD), and RMSE.*

	Auditory			**Visual**		
	Mean	**SD**	**RMSE**	**Mean**	**SD**	**RMSE**
Level 1	0.1208	0.0568	14.264	0.1325	0.0428	15.535
Level 2	0.1219	0.0657	15.431	0.1413	0.0451	16.554

Subsequently, an independent two sample t-test was conducted with the absolute accuracy between conditions and levels. However, any effect was found between the auditory and visual conditions in Level 1 [t(21)=-0.5613, p-value=0.5805] or Level 2 [t(21)=-0.8305, p-value=0.4156]. Furthermore, any effect was found between Level 1 and Level 2 with the auditory condition [t(20)=-0.0430, p-value=0.9662] or with the visual condition [t(22)=-0.4901, p-value=0.6289]. Therefore, the analysis considered running a linear mixed regression model (mixed models) for absolute accuracy.

Linear mixed-effects models (LMMs) are increasingly being used for data analysis in cognitive neuroscience and experimental psychology, where within-participant designs are common (Magezi, 2015). LMM describes the relationship between a response variable and other explanatory variables that have been obtained along with the response (Magezi, 2015). For instance, in some experiments, one participant provides multiple data points (e.g., tapping). Those data will correlate with one another because they come from the same participant (Shrikanth, nd). Explanatory variables can be categorized as “fixed factors” or “random factors” (Gagnon et al., nd; Magezi, 2015). Fixed effects are those where all interest levels are included in the experiment. The fixed effect is related to data collected from all levels (e.g., gender: male/female/other; condition: auditory/visual; complexity: Level 1/Level 2, age, number of events). Meanwhile, the random effect includes only a sample of all possible levels. Even though most of the studies include a large population, such as adult humans, psychology investigations generally only include a tiny subset of that population (e.g., participants) (Magezi, 2015). Thus, the ability to include various configurations of grouping hierarchies (multiple, nested groups such as street, town, country, and continent) is advantageous for social scientists. Furthermore, LMM results provide better interpretability regarding physiological phenomena and a superior fit to the data (Magezi, 2015).In this study, the variable of interest is the accuracy of the participants. This accuracy is related to different stimuli (auditory or visual, between-subjects), visual guide support (game-elements, within-subjects), and variables such as gender and age.

Nevertheless, each subject has a different time reaction which will be an idiosyncratic factor that affects all responses from the same subject (reaction time of each tap and number of taps). To render these different responses inter-dependent (correlated) rather than independent, the model considers a different mean accuracy for each participant. Thus the model assumes different random intercepts for each participant (Gagnon et al., nd). Therefore, the formula for the model for this study is

$$Accuracy = Intercept + No.Beat + Level + Condition + Age + Gender + (1 \mid Participant) \tag{3.2}$$

Where *int* is the random intercepts, *No.Beat* is the event number of the rhythm sequences vector. *Level* refers to Level 1 or Level 2. *Condition* refers to Auditory or Visual stimuli. The range for *Age* is 17 - 41 years old. *Gender* considers female, male, or other.

Table 3.2 shows the results of the linear mixed-effect model for accuracy. The number of beats at which the measurement was recorded had a significant association with the accuracy (p-value = 0.019). The effect estimate of 0.0003s suggest a fatigue effect of 0.0003 seconds per tap, corrected for all other variables. Associations between target variables and accuracy were not found.

TABLE 3.2: *Results from the linear mixed regression for accuracy, bold font highlights significant effects.*

Effect	Estimate	Standard Error	t Value	Pr >|t|	Lower 95% CL	Upper 95% CL
Intercept	0.131	0.029	4.564	0.0005	0.074	0.187
Visual vs. Auditory	0.014	0.026	0.547	0.584	-0.036	0.064
Beat No.	**0.0003**	**0.0001**	**2.339**	**0.019**	**0.00005**	**0.0006**
Level 2 vs. Level 1	0.008	0.005	1.696	0.089	-0.001	0.019
Male vs. Female	-0.037	0.030	-1.227	0.219	-0.096	0.022
Non-Musician vs. Musician	0.025	0.023	1.083	0.278	-0.021	0.072

Results for Task Load measured by NASA TLX can be seen in Table 3.3. The difference between scores of 40.60 (auditory) and 40.35 (visual) was not significant for Level 1; meanwhile, for Level 2, the difference between auditory condition (46.25) and visual condition (49.72) was also not significant.

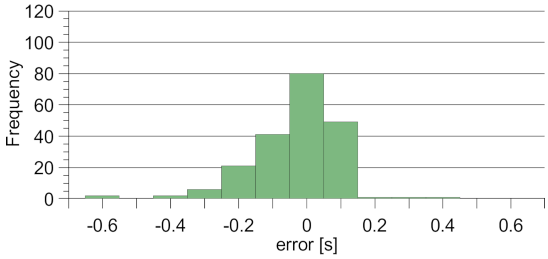
TABLE 3.3: *Results of NASA TLX ([0;100], the lower the better) (mean ± standard deviation (SD)).*

	Auditory	Visual
Level 1	40.69 ± 19.77	40.35 ± 18.11
Level 2	46.25 ± 16.17	49.72 ± 19.42

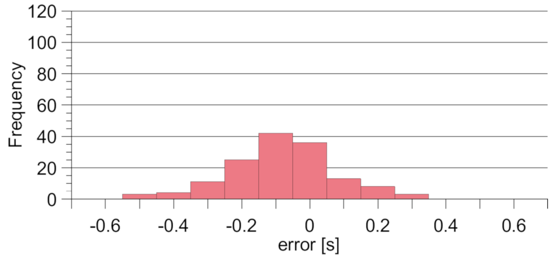
3.3.4 Results Considering the Music Background of Participants

On one hand, an analysis taking into account the musical background from participants was performed. As aforementioned, there were six musicians and six non-musicians in each condition, auditory and visual. Therefore, data were divided into musicians-auditory condition, non-musicians-auditory condition, musicians-visual condition, and non-musicians-visual condition.

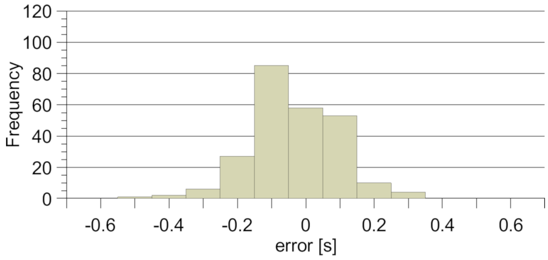
Therefore, to identify where the variation is, the data were represented using histograms. Moreover, histograms show whether the participants intended to anticipate



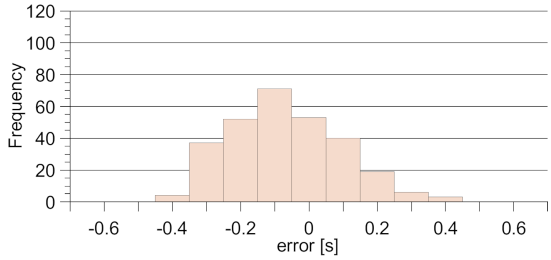
(A) *Level 1 Musician - Auditory*



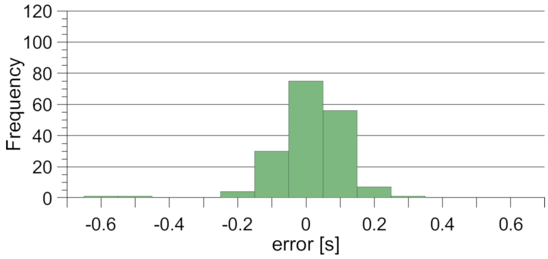
(B) *Level 1 Non-Musician - Auditory*



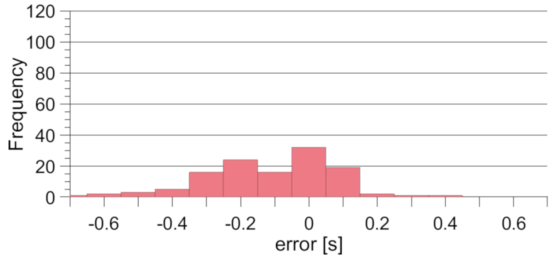
(C) *Level 1 Musician - Visual*



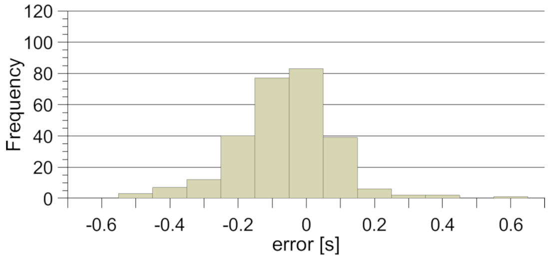
(D) *Level 1 Non-Musician - Visual*



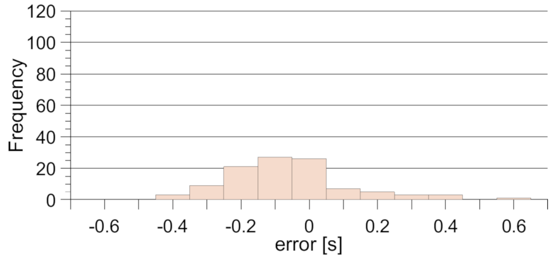
(E) *Level 2 Musician - Auditory*



(F) *Level 2 Non-Musician - Auditory*



(G) *Level 2 Musician - Visual*



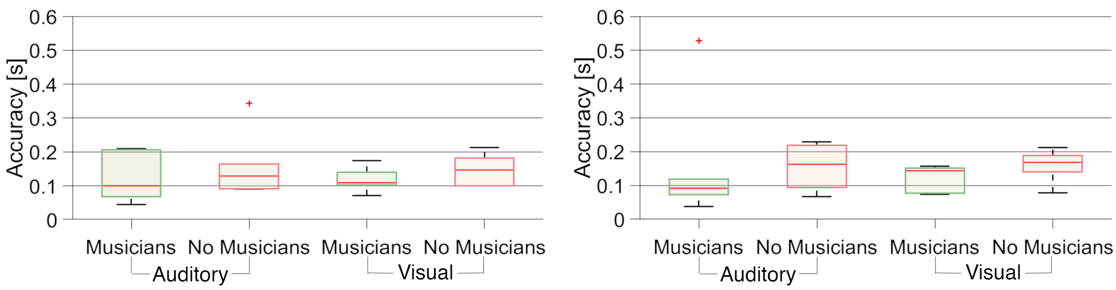
(H) *Level 2 Non-Musician - Visual*

FIGURE 3.11: *Histograms of auditory and visual conditions in Level 1 and Level 2 divided by musicians and non-musicians. The histograms show the distribution of data of all the participants in each condition. The positive values indicate that the player pressed the space bar of the computer keyboard before the stimuli appears. Negative values indicate that the player pressed after the stimuli, which indicates a reaction.*

the stimuli (negative values) or tap after the stimuli (positive values). Fig. 3.11a shows that most of the musician-auditory taps were around 0s in Level 1. In contrast, non-musicians had fewer taps, and most of them nearby -0.1s (Fig.3.11b). On the same level but for visual condition, musicians had more taps around -0.1s (Fig.3.11c) than non-musicians 3.11d. However, histograms display a tendency of participants under the visual condition to anticipate the stimuli.

For Level 2, musicians with auditory condition performed most of the taps around 0s (Fig.3.11e). In contrast, taps of non-musicians were not consistent, so there is not a specific range (Fig.3.11f). Results of musicians with the visual condition display that most of the taps of musicians were around 0s (Fig.3.11g), and also a higher number of taps were in the range of -0.1s. Meanwhile, taps of non-musicians with visual conditions were not enough to show a precise histogram (Fig.3.11h).

Similar to section 3.3.3, anticipation or delays were not the value of interest; thus, the analysis was focused on the accuracy as a magnitude of time. For this analysis, a boxplot of the absolute accuracy shows the distribution per level, condition, and music background. The results for Level 1 auditory condition, showed a median of 0.0994s for musicians and a median of 0.1288s for non-musicians (Fig.3.12a). Regarding visual condition, musicians obtained a median of 0.1088s, whereas the median for non-musicians was 0.1458s. An outlier of 0.3442s was found for musicians with the auditory condi-



(A) *Level 1* (B) *Level 2*

FIGURE 3.12: *Boxplot of musicians and non-musicians in auditory and visual condition at Level 1 and Level 2. Boxplot shows the dispersion of the absolute value of data from the participants, based on the minimum and maximum values, median, first and third quartile, and possible outliers.*

tion. In Level 2, the auditory condition displays a median of 0.0912s for musicians and 0.1622s for non-musicians (Fig.3.12b). Results show a second outlier of 0.5281s for musicians with the auditory condition. For visual condition, musicians had a median of 0.1426s, while for non-musicians, it was of 0.1675s.

As done previously, the Root Mean Square Error (RMSE) was calculated to identify in which condition participants performed closer to a perfect synchronization. Using equation 3.1, for Level 1, the auditory condition for musicians displays an RMSE of 13.0994. In comparison, RMSE for non-musicians was 15.8515.
On one hand, for visual condition, RMS was of 13.3175 for musicians and 17.2210 for non-musicians (Table 3.4). Meanwhile, for Level 2, the RMSE for the auditory condition was 9.698 for musicians and 21.6445 for non-musicians. Regarding visual condition, RMSE was 15.4783 for musicians and 19.0592 for non-musicians (Table 3.4).

TABLE 3.4: *Results of Mean, Standard Deviation (SD), and RMSE considering the music background of participants*

		Auditory			Visual		
		Mean	**SD**	**RMSE**	**Mean**	**SD**	**RMSE**
Level 1	**Musicians**	0.1212	0.0721	13.099	0.1173	0.0354	13.317
	Non-musicians	0.1203	0.0394	15.851	0.1476	0.0472	17.221
Level 2	**Musicians**	0.0821	0.0305	9.698	0.1237	0.0382	15.478
	Non-musicians	0.1551	0.0707	21.644	0.1588	0.0476	19.059

Similarly, absolute mean accuracy in seconds was computed per subject and condition, considering participants' background. Therefore, for Level 1, musicians show a mean of 0.1212s with a standard deviation of 0.0721s for the auditory condition. At the same time, the non-musicians display a mean of 0.1203s with a standard deviation of 0.0394s. For visual condition, musicians had a mean of 0.1173s and a standard deviation of 0.0354s, and non-musicians displayed a mean of 0.1476s with a standard deviation of 0.0472s (Table 3.4). Concerning Level 2, the mean for musicians with the auditory condition was of 0.0821s and a standard deviation of 0.0305s. In contrast, for non-musicians, the mean was of 0.1551s, with a standard deviation of 0.0707s. Regarding the visual condition, musicians had a mean of 0.1237s and a standard deviation of 0.0382s. In comparison, the mean for non-musicians was of 0.1588s, with a standard deviation of 0.0476s (Table 3.4).

This section aimed to determine whether participants' music background affects their accuracy (reaction time) using auditory condition or whether music background affects participants' accuracy using visual condition. A t-test was carried out on the absolute accuracy between conditions and levels. However, for the auditory condition, not found was any effect of music background in Level 1 [t(9)=0.0240, p-value=0.9814] or Level 2 [t(9)=-2.1338, p-value=0.0616]. Meanwhile, for visual condition, it was not found any effect of music background in Level 1 [t(10)=-1.2587, p-value=0.2367] or Level 2 [t(10)=-1.4101, p-value=0.1888]. In consequence, the analysis considered running a linear mixed regression model for absolute accuracy.

For this analysis, the linear mixed-effect model considered not only variables such as auditory or visual condition, beat number, level, age, and gender; but this time model also analyzed whether a correlation exists between participants' condition and background. Therefore, the formula for the model for this study is

$$\begin{aligned} Accuracy = Intercept + No.Beat + Level + Condition + Age + Gender + \\ Condition * Background + (1 \mid Subject) \end{aligned} \tag{3.3}$$

Where *int* is the random intercepts, and *No.Beat* is the event number of the rhythm sequences vector. *Level* refers to Level 1 or Level 2. *Condition* refers to Auditory or Visual stimuli. The range for *Age* is 17 - 41 years old. *Gender* considers female, male, or other.

Considering participants as a random effect, the results of the model are shown in Table 3.5. Results did not change when the analysis involved the correlation between condition and background of participants. Therefore, the number of beats at which the measurement was recorded had a significant association with the accuracy (p-value = 0.018).

TABLE 3.5: *Results from the linear mixed regression for accuracy with a possible effect of the background into the condition, bold font highlights significant effects.*

Effect	Estimate	Standard Error	t Value	Pr >|t|	Lower 95% CL	Upper 95% CL
Intercept	0.133	0.032	4.154	0.0003	0.070	0.196
Visual vs. Auditory	0.009	0.036	0.266	0.790	-0.061	0.080
Beat No.	**0.0003**	**0.0001**	**2.355**	**0.018**	**0.00005**	**0.0006**
Level 2 vs. Level 1	0.008	0.005	1.702	0.089	-0.001	0.019
Male vs. Female	-0.036	0.031	-1.178	0.238	-0.098	0.024
No-Musician vs. Musician	0.021	0.035	0.616	0.537	-0.047	0.090
Visual/Auditory & No-Musician/ Musician	0.008	0.048	0.166	0.868	-0.087	0.103

The effect estimate of 0.0003s suggest a fatigue effect of 0.0003s per tap. No more associations between the target variables and accuracy were found.

3.3.5 Discussion

The goal of the evaluation mentioned above was to analyze the performance of participants playing a video game through two modalities: auditory and visual. In addition, the case study aimed to investigate how game-elements affect the execution. Participants played the game *Jump'n'Rhythm* (Alexandrovsky et al., 2016) in two levels of difficulty. Level 1 showed key game-elements (i.e., platforms and gaps) and one of the modalities; whereas Level 2 hid these key game-elements but kept the rest of the game's dynamic similar to Level 1. Evidence of the effects of synchronization using auditory or visual stimuli in the video game environment was gathered.

Contrary to expectations, this study did not find a significant difference between auditory or visual stimuli. Moreover, the presence of supportive game-elements (Level

1) or their absence (Level 2) was not statistically significant. Therefore, hypothesis H1 (*Players have a better reaction time using auditory stimuli even in the absence of supportive game-elements*) was not confirmed. However, some insights from the observations of the data during the analysis can be expressed. During the process, the RMSE was calculated to determine how far or close the data were from perfect accuracy (0s). In Table 3.1, the auditory condition had an RMSE closer to 1 (the data fit totally with the model) for both levels. Therefore, participants' time difference values using the auditory condition were closer than the participants using the visual condition to 0s. As mentioned in the literature review, the synchronization task is better using auditory stimuli rather than visual (Repp and Penel, 2002; Chen et al., 2002; Repp and Penel, 2004; Patel et al., 2005; Hove et al., 2013a,b). Moreover, during the case study, participants manifested more confidence in performing the task following auditory rather than visual stimuli. Regarding the use of supportive game-elements, Level 1 had a lower value of RMSE for both conditions than Level 2. A possible explanation for this is that in Level 1 participants followed the game-elements and not the corresponding stimuli. Meanwhile, in Level 2, participants did not have the game-elements to support their performance. They depended only on the stimuli. However, without significant differences, caution must be applied to these assumptions. Another interesting effect (in both conditions and levels) that was not analyzed but observed during the data processing, was the participants' tendency to react to stimuli instead of anticipating. However, in the case of participants with the auditory condition, the effect was more evident (Fig.3.9).

Surprisingly an association between the accuracy and the number of patterns performed by participants was found (Table 3.2). Each time the participant moves forward by pressing the space bar, the participant had a delay. This delay was 0.003s (p-value=0.019) per tap. This association might result from a fatigue effect during their performance.

Regarding H2 (*Musicians have a better reaction time than non-musicians using auditory or visual stimuli*), the results of this study indicate that no statistical significance was found between the music background and conditions. Therefore, the hypothesis was not confirmed. However, from the data observation, according to the RMSE, musicians had better accuracy with auditory and visual conditions at both levels than non-musicians (Table 3.5). Contrary to expectations, the non-musicians in Level 2 had a lower RMSE with the visual condition than with the auditory. This result may be explained by the fact that the challenge and competence of Level 2, in addition to visual conditions, provoke more attention by participants. This result is not far from expected results because musicians follow auditory rhythm patterns naturally. However, these results need to be interpreted with caution because they are not statistically significant. From the observation of the data distribution, most of the taps of the musicians were around 0s in comparison with the non-musicians whose taps were less precise (Fig. 3.11). Moreover, for either musicians or non-musicians there was a tendency to react to the stimulus and not to anticipate the stimulus.

One unanticipated finding was that an association between the accuracy and the number of patterns performed by participants was statistically significant (Table 3.5). Each time a participant moves forward by pressing the space bar, the participant had a delay. This delay was 0.003s (p-value=0.018) per tap. A possible explanation for this might be that this association might result from the fatigue effect during their performance.

The results of the Nasa-TLX were not statistically significant; in consequence, hypothesis H3 (*The workload of participants using visual stimuli is higher than participants using auditory stimuli*) was not confirmed. However, in Table 3.3, TLX results for Level 1 are not far from each other with the auditory or visual condition. Nevertheless, the visual condition had a slightly lower value than the auditory condition. The reason for this is not clear, but it may have something to do with the stimuli. Participants with the visual condition most likely followed only the supportive game-elements, and they ignored the stimuli. In contrast, the auditory condition in a temporal task can affect the performance; thus, players could be confused about whether to follow either the auditory stimuli or the game-elements (Bootsma and C. W. Van Wieringen, 1990; Tresilian, 1994; Miyake et al., 2004; Hove et al., 2013b). These findings may be somewhat limited because no statistical significant was found.

While preliminary, these observations suggest that the accuracy of players using auditory stimuli is better than visual stimuli in the presence or absence of key game-elements through a video game environment. Thus, according to literature, auditory modality improves rhythm skills (Repp and Penel, 2004; Leman, 2008; Kylie et al., 2011; de Castell et al., 2014). Another aspect is the presence of key game-elements that probably influence the perception of the stimuli. Thereby, the number of key game-elements that contain rhythm games is vital to improving rhythm skills. Otherwise, players follow game-elements and they do not train their auditory-motor system, which is essential for rhythm learning.

These findings will doubtedly be scrutinized, but some immediately dependable conclusions are drawn for possible fatigue effects. It could be necessary to consider long lapses for resting between sequences to avoid the fatigue effect after playing the game a few times.

There are still many unanswered questions about the influence of a rhythm video game's auditory or visual stimuli in the players' behavior. Thus in future, it is necessary to consider a long-term study using auditory modality to measure the improvement of rhythm skills in participants. Moreover, some evaluation using behavioral psychology methods to find the connection between the players' behavior and perception and the video game is necessary. Further studies, which take these variables into account will need to be undertaken.

Regarding the prototype design, an improvement in the game with more levels, tempo, and more rhythms is possible. As mentioned, rhythm is a fundamental element for reading activities; therefore, a new version of the game is in process. This version of the prototype intends to conduct a study to detect developmental coordination disorders.

3.4 Auditory and Tactile Stimuli in Sychronization Tasks

As previously mentioned, rhythm is a fundamental part of daily life and is related to the different internal processes of the body. The synchronization between body movements and a sequence of sensory events or patterns is essential to acquire rhythm skills. This ability is related to the human sensorimotor system (Lahav et al., 2005; Hove et al., 2013b), which is involved when people begin moving part of their body while listening to music or rhythm. However, the body acquires rhythm through auditory stimuli, and people can also "feel" the music through the somatosensory system (Huang et al., 2013). Mechanoreceptors cells (Pacinian) in the skin are activated when detecting vibration; then, they can encode temporal patterns as sound. In video games, some game console controllers often incorporate tactile output. Therefore, tactile stimuli should be explored as an alternative for rhythm skills training using video games. The present study analyzes the influence of auditory or tactile stimuli in players. Therefore, the game *Jump'n'Rhythm* was adapted to analyze players' performance in a rhythm video game environment with two modalities: auditory or tactile. For this study, three hypotheses were stated:

- H4: The performance of the players using tactile stimuli is similar to the performance of players using auditory stimuli.
- H5: The absence of game-elements hinders the performance of players in both conditions.
- H6: The user experience is higher for participants using the auditory condition.

3.4.1 Case Study

The case study examined the reaction times of players following either auditory or tactile stimuli in a video game environment. The game *Jump'n'Rhythm*, described in section 3.2, was used to evaluate the performance of players in both conditions, auditory or tactile. Participants with the tactile condition used a vibrotactile device on their body, especially designed for this study. The study consisted of two groups of participants; one group followed the auditory stimuli and another group, tactile stimuli.

Participants. Twenty-seven participants (seven females and twenty males) between twenty to thirty-one years of age took part in the study. All participants live around Bremen area (Germany), and most of them, twenty-four, belong to the University of Bremen. All participants reported that they had normal hearing and tactile sensation. From this group of participants, three participants are left-handed, and one is ambidextrous; the rest of the participants are right-handed. The participants were split into two: a group in the auditory condition and a group in the tactile condition. The group in the auditory condition had fourteen participants, where three of them were women. On the one hand, the group in the tactile condition had thirteen participants, where four of them were women. Seven of the participants are actively playing an instrument. Five of them have played for more than four years already; however, none of them is a professional

Apparatus. The participants played the game *Jump'n'Rhythm* (Alexandrovsky et al., 2016) running in Unity[11] on a Lenovo Thinkpad E540. A 17-inch Hanns-G external monitor with a resolution of 1440x900px displayed the game, which had a resolution of 1368x770px. The auditory stimuli were transmitted to the participant in the auditory condition by AKG k271 Studio headphones. Moreover, headphones supported participants in the tactile condition to shield them from possible external noise. For the tapping task, participants used an external mechanical computer keyboard. The volume of the signal was set to an appropriate magnitude so that the signal was the same for all participants, and they could hear stimulus.

The Stimuli. The stimuli were sequences presented with a tempo of 90bpm (one beat every 0.667s); thus, players could perform the synchronization task comfortably. Metrical sequences in the prototype include intervals of only 0.667s, considering quarters

[11] Unity Technologies. (2005). *Unity*. [Game Engine]. Available in `https://unity.com/`

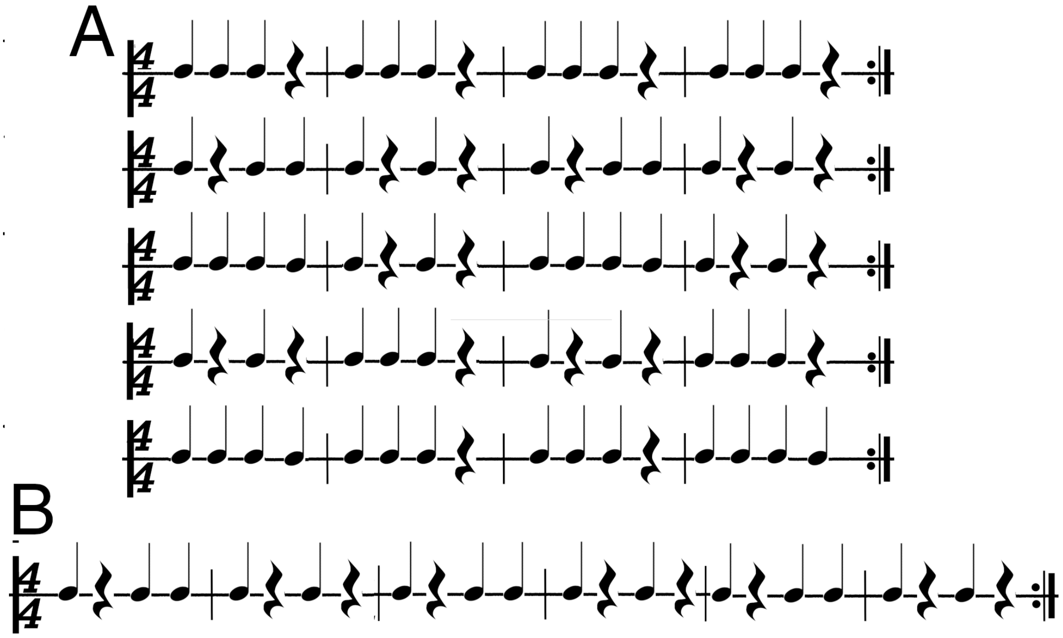


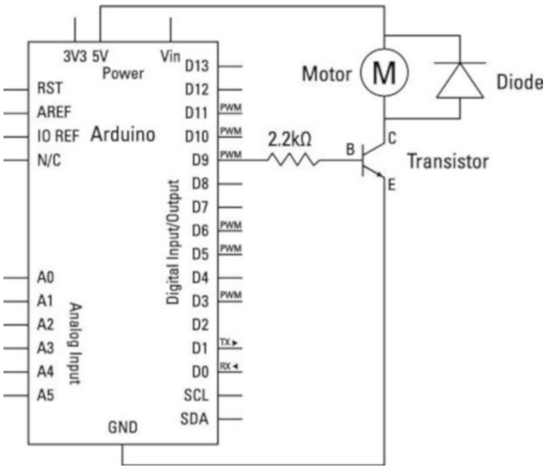
FIGURE 3.13: *Rhythm patterns or sequences emitted by the video game Jump'n'Rhythm (Alexandrovsky et al., 2016).* ***A)*** *The five sequences which are executed by the game, in order from top to bottom.* ***B)*** *Sequence used in the pre test and post test. Each quarter note activates the stimulus, either auditory or tactile; therefore, the stimuli follow the rhythm sequence.*

and silences (Fig.3.13) in a music context.

In the auditory condition, the stimulus for rhythm patterns is an audible tone of 1kHz emitted through the video game. Tone uses a percussive envelope consisting of an attack time of 0.005s and a release time of 0.110s. Each stimulus appears according to the rhythm sequences in Fig. 3.13-A.

The tactile stimulus consisted of a vibration (240Hz) with a duration of 0.30s, which follow rhythm sequences in Fig 3.13. The motor to produce these vibrations is a shaftless vibration motor of 8x3.4mm with a vibration amplitude of 3V 0.75g embedded on an elastic fabric wristband (Fig. 3.14b). An Arduino Duemilanove board[12] controlled the motor via the serial port, according to the diagram of Fig. 3.14a.

[12] Ardunio.cc.(2009). Arduino - Arduino Duemilanove: `https://www.arduino.cc/en/Main/arduinoBoardDuemilanove` [Accessed 20 June 2019].



(A) *Diagram of connexions used to connect the different electronic components.*

(B) *The motor was placed inside the elastic fabric wristband and attached to the forearm of the players.*

FIGURE 3.14: *The figure displays the elements used to build the prototype.*

Since the motor required only 120mA max and operating voltage was in the range of 2.5 – 3.5V, it was not necessary to use an external battery for the motor; thus, electric power was distributed by Arduino[12]. A transistor 2N2222 was selected because it offers a collector current of 800mA that is double the stall current (120mA) of the motor. A diode protected the circuit to returning electromotive forces. Intensities of tactile stimuli were set to a level, which produced a sufficient alerting effect to create responses from participants. The device position was on the forearm of participants, as its sensitivity and ability to distinguish signals is nearly as good as the fingertip, and because the tapping task was with the fingertip (Jones and Sarter, 2008).

Procedure. The experiment was conducted at the University of Bremen. The experiment consisted of only one session in a quiet room with good lighting, where participants remained seated in front of a monitor and computer keyboard. The experiment was a between-subjects design, where participants performed only one of the two conditions, auditory or tactile. At the beginning of the session, participants completed a consent form and were informed about the goals of the study.

The next step of the session was a pre test, which consisted of two phases: beat tapping and sequence processing. In the beat tapping phase, participants tried to synchronize their taps in a sequence (Fig.3.13-B). Afterwards, in the sequence processing phase, the same sequence was reproduced, and once the sequence finished, participants were required to tap the same sequence without any stimuli. Auditory or tactile stimuli emitted rhythm sequences. The test was performed using a totally black screen, and a beep at the beginning indicated that the test had begun. This test was done once before and once after (post test) participants played the game.

In the last part of the session, participants completed three questionnaires: 1) unweighted NASA task load index (TLX) (Hart and Staveland, 1988), 2) Usability Metric for User Experience Lite (UMUX-Lite) (Lewis et al., 2015), and 3) Intrinsic Motivation Inventory (IMI) (McAuley et al., 1989). The UMUX is a short version of the System Usability Scale. IMI assesses dimensions of interest-enjoyment, effort importance, competence, and pressure-tension during play. Finally, participants completed a demographic questionnaire that asked basic questions about the players' expertise in gaming and music.

Data Processing. Questionnaires UMUX-Lite, NASA TLX, IMI registered the data of each participant. Similarly, data of the reaction time of players were acquired for each participant. For the acquisition of these data, a log file was created which was embedded inside the game *Jump'n'Rhythm*. The log file created an ASCII file to register when a player pressed the space bar on the computer keyboard; every time the avatar died, a new ASCII file was created. Simultaneously, the log file creates another ASCCI file to register the time when each event of the rhythm sequence occurred, either with auditory or tactile sequences. Similarly, a new ASCII file was generated every time that the avatar died. The files were exported to Matlab[13] to continue with the analysis. The first part was to normalize the length of the time series by padding them with zeros in the time-domain. Afterward, the time series for taps and the time series for sequence events were plotted. The intervals of the rhythm sequences were defined manually.

[13]MATLAB 2018a, The MathWorks, Inc., Natick, Massachusetts, United States

Each interval contained a tap and an event of the stimuli (auditory or tactile). Subsequently, the time differences between the tap and the stimuli event were calculated for each player and level, considering the number of lives the player used. In the end, there was a set or vector of time differences for each player. These vectors were analyzed using SAS[14] software version 9.4 to find correlations between level, condition, age, gender, and participants.

For pre-test and post-test, the study expected that participants tap according to the stimuli events and the required task. However, most of the participants failed in the tasks by omitting or adding taps to the rhythm sequences. A penalty for each mistake was added to take into account these mistakes and grade them. Therefore, when the participant did not tap during a stimulus event, a penalty of 0.9s was added for that mistake. In contrast, when the participant tapped the space bar and no stimulus event was in the rhythm sequence, a penalty of 0.9s[15] was added for that mistake. As a result, the time difference vector will have extreme values of 0.9s, which means that the participant's accuracy (performance) was not optimal. To obtain this time difference vector, the log file generates two ASCII files, which were processed and analyzed, as mentioned previously.

Data Calibration. In the previous case study, where the auditory and visual conditions were analyzed, the data's acquisition was audio recordings. This method was similar to those used in the experiments of Tierney et al. (2017), and was also commented on in section 3.3.1. However, one of the aims of this case study was to generate a method to obtain reaction times within the game. Therefore, the log file was implemented and evaluated by comparing the audio recording method (section 3.3.1) and the log file method to validate the log file's efficiency. A simple test compared the two methods. For this test, the game was played in two rhythm sequences using auditory conditions. The data were registered by audio recording method and by log file

[14]Copyright © 2013 SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA

[15]The value of 0.9s was established because it is a value far away from the perfect accuracy (0s)

method. On one hand, audio recording files were edited following the same process described in section 3.3.1 to obtain the time differences. On the other hand, ASCII files generated by the log file were processed following the section described before. Each of the two methods obtained their corresponding time differences vector. Afterward, the two vectors were compared to corroborate whether the data registered by the log file behave similarly to the validated method of Tierney et al. (2017). As shown in Fig.3.15, a line was drawn to visualize a trajectory by joining the time difference (dots). These trajectories were similar, either for the log file or audio recording method.

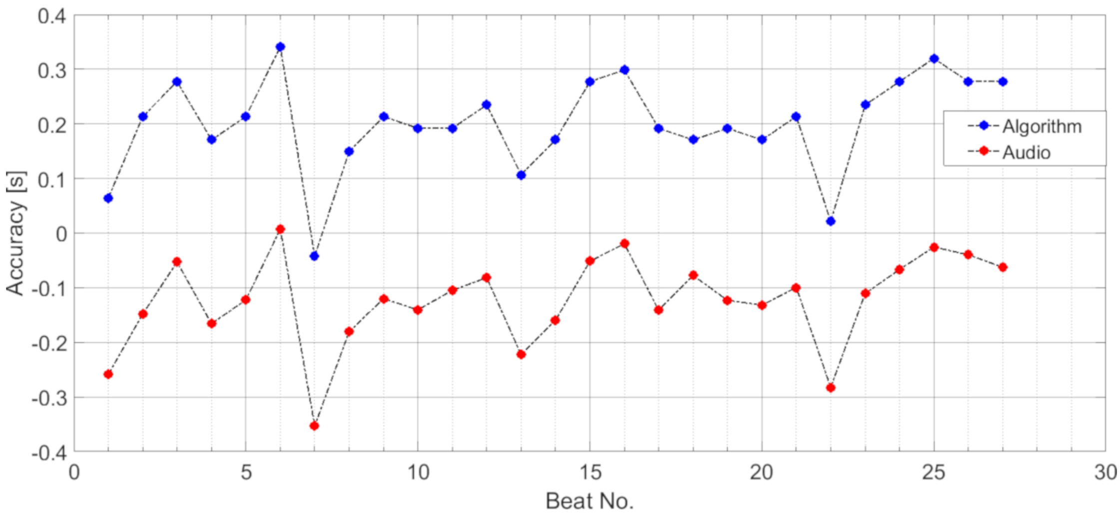


FIGURE 3.15: *This figure shows the two data sets of the time differences between the stimuli and tapping. Blue dots are the data obtained from the algorithm; each blue dot represents the accuracy for the corresponding beat. Red dots indicate the data obtained from the audio record and represent the accuracy of that beat. The dashed black line is symbolic, and it is drawn to join the dots to create a shape between dots. Thus, it is clearer to look at the similarities between the two data sets. For instance, the red dot from the audio record in the beat five indicates an accuracy of -0.13s approximately, and its corresponding fifth beat in the blue dot indicates an accuracy of 0.21s using the algorithm.*

To quantify the degree of similarity between the two data sets of time differences, the cross-correlation was calculated. Cross-correlation functions provide a measure of association between two different signals (Derrick and Thomas, 2004; Sneha, 2017) and it is the tool most commonly used in the analysis of multiple time series (Peterson, 2001). The cross-correlation between two vectors or time series is given by the function (Orfanidis, 1988)

$$\mathbf{R}_{xy}(m) = \mathbf{E}\{x_{n+m}\, y_n^*\} = \mathbf{E}\{x_n\, y_{n\text{-}m}^*\} \tag{3.4}$$

where x_n and y_n are the two data set or vectors of time differences, n is an integer $-\infty < n < \infty$, and $\mathbf{E}\{\}$ is the expected value operator (Northwestern University, 2017).

This function is already implemented in Matlab as xcorr(x,y) where calculates the cross-correlation of two discrete-time sequences. Whether x and y have different lengths, the function appends zeros to the end of the shorter vector to have the same length as the other. Moreover, when the lag is not specified as an integer scalar, the lag range equals $2N$–1, where N is the greater of the lengths of x and y. Therefore, the cross-correlation between the log file vector and the audio recording method is shown in Fig. 3.16.

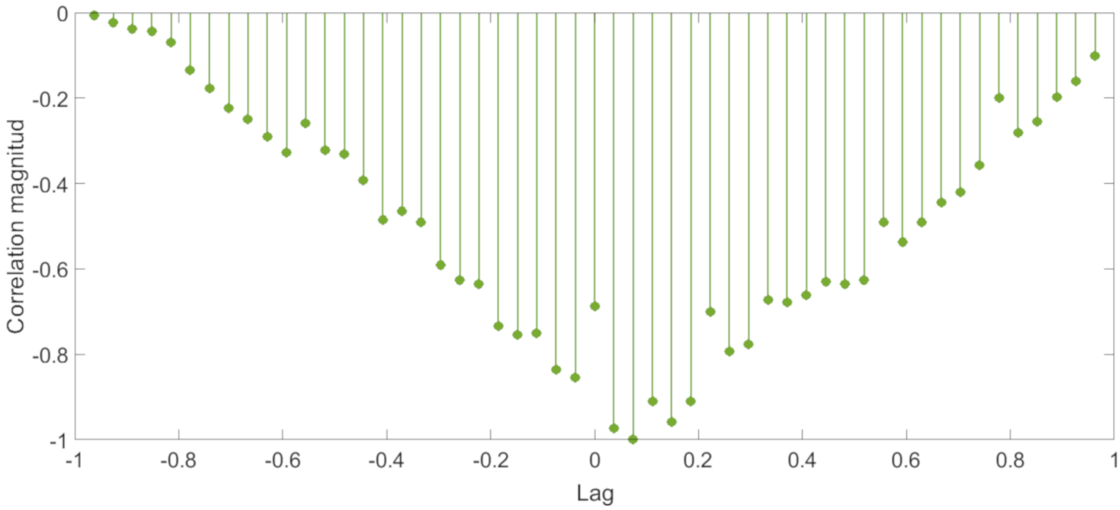


FIGURE 3.16: *This figure displays the cross-correlation between the time difference's audio recording vector and the log file vector of time differences. The highest peak of the cross-correlation signal occurs at the lag where the elements of both signals match exactly (0.07). When the highest peak value is zero or closer, both signals are similar in most of their points. The stems represent the cross-correlation magnitude (or how similar the vectors are on that lag "point") for that specific lag. As the length of the time difference vectors is 27 values, the lag range is 53 points (in the figure, 53 stems). X and Y axes are normalized.*

Consequently, knowing that the time series vectors obtained from the audio recording method and the log file method are similar, the log file method is valid to acquire the data of the case study.

3.4.2 Results

Data measurements using the algorithm allows identifying the number of beats progressed and the number of lives per player during each level. Using this information, data were analyzed by levels and conditions. Afterward, analyses considered the intersection between condition, number of beats, and levels. According to Fig. 3.17, for the auditory condition, 50% of participants achieved a total of beats (58 beats) and finished the level using only one life. Whereas 21.43% of participants used two lives, the rest (28.57%) used a third life. In comparison, with the tactile condition, 30.77% of participants finished the level in the first trial, 15.38% of participants used two lives, and more than half of the participants (53.85%) used three lives.

Fig.3.17 shows a lack in the performance of players during Level Level 2 in comparison to Level 1. Using auditory condition, only two participants finished the level: one participant used one life, and the other participant used two lives. The rest of the par-

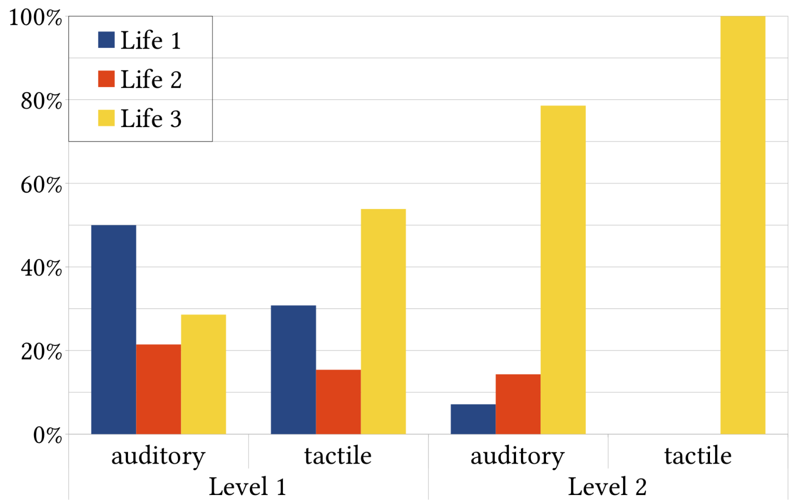


FIGURE 3.17: *Overall results expressed in percentage of the number of lives used by the participants per level and stimulus.*

ticipants (85.71%) used three lives. In contrast, all participants with the tactile condition used three lives.

The value of interest was accuracy in terms of time. Thus, the analysis considered absolute values (of the time differences) of the data also called absolute accuracy. The first part of the analysis was to verify the data distribution of the absolute accuracy by plotting a boxplot. The boxplot helps determine where the distributions lie concerning one another by plotting the quantiles side by side.

Therefore, the boxplot results for Level 1 show absolute data distribution with a median of 0.23s for the auditory condition and a median of 0.26s for the tactile condition. In contrast, in Level 2 for the auditory condition, the median was 0.20s and a median of 0.25s for the tactile condition. Any outliers were found (Fig.3.18).

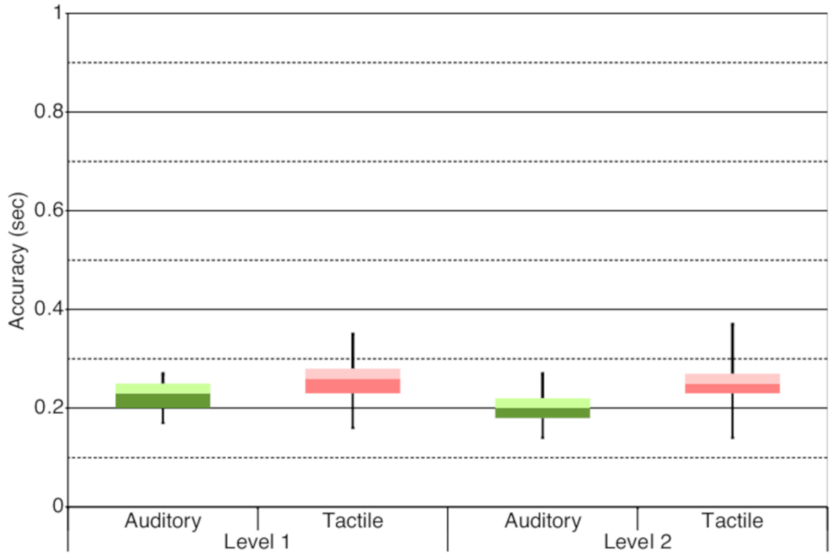


FIGURE 3.18: *Boxplot of auditory and tactile condition at both levels. The boxplot shows the accuracy of the participants.*

Subsequently, absolute mean accuracy and the standard deviation were computed for both levels and conditions. Therefore, for Level 1, the auditory condition displays a mean of 0.22s and a standard deviation of 0.03s. In contrast, the tactile condition shows a mean of 0.25s and a standard deviation of 0.05s. Level 2 presents a mean of 0.20s and a standard deviation of 0.04s for the auditory condition. In comparison, the tactile condition had a mean of 0.25s and a standard deviation of 0.06s.

Afterward, two sample t-test was computed of the absolute accuracy between conditions and levels. However, not found was any effect on the auditory and tactile conditions in Level 1 [t(25)=-.1.6861, p=0.1042]; nevertheless, in Level 2, it was found [t(25)=-2.2972, p=0.03]. Moreover, it was not found any effect of either Level 1 and Level 2 with the auditory condition [t(26)=1.6167, p=0.1180] or visual condition [t(24)=0.2889, p=0.7751]. Nevertheless, this test was inconclusive; therefore, the data were analyzed running a linear mixed regression model for absolute accuracy.

As it was explained on page 64 and 65, mixed models work around the assumption that data points (taps) are independent of one another. The model is useful when participants provide multiple data points. Moreover, using a mixed model accounts for variation due to the independent variables (e.g., gender: male/female/other; condition: auditory/visual; complexity: Level 1/Level 2; age, number of events) and variation that is not explained by the independent variables of interest (participants).
Therefore, the formula for the model for this study is

$$Accuracy = Intercept + No.Beat + Level + Condition + Age + Gender + (1 \mid Participant) \tag{3.5}$$

Where *int* is the random intercepts, *No.Beat* is the event number of the rhythm sequences vector. *Level* refers to Level 1 or Level 2. *Condition* refers to Auditory or Tactile stimuli. The range for *Age* is twenty to thirty-one years old. *Gender* considers female, male, or other. The results of the model are shown in Table 3.6.

Table 3.6 shows results of the model with a significant difference between auditory and tactile stimuli concerning accuracy. Participants under the tactile condition reached an absolute deviation from optimal point 0.037s higher than participants with the auditory stimuli (p-value = 0.0177). Furthermore, the analysis shows a delay of 0.001s (p-value = 0.00007) per jump as the game progresses, most likely associated with a fatigue effect. A better accuracy was found in life 2 (-0.028, p-value = 0.0001) and life 3 (-0.032, p-value = 0.0002) than in life 1. Nevertheless, life 2 and life 3 did not reveal a

TABLE 3.6: *Results from the linear mixed regression for accuracy, bold font highlights significant effects.*

Effect	Estimate	Standard Error	t Value	Pr >|t|	Lower 95% CL	Upper 95% CL
Intercept	0.261	0.069	3.78	0.001	0.118	0.403
Tactile vs. Auditory	**0.037**	**0.016**	**2.37**	**0.0177**	**0.007**	**0.068**
Beat No.	**0.001**	**0.000**	**3.39**	**0.0007**	**0.000**	**0.001**
Level 2 vs. Level 1	-0.005	0.006	-0.83	0.4076	-0.016	0.007
life 2 vs. life1	**-0.028**	**0.007**	**-3.83**	**0.0001**	**-0.043**	**-0.014**
life 3 vs. life1	**-0.032**	**0.009**	**-3.71**	**0.0002**	**-0.049**	**-0.015**
Age	-0.001	0.003	-0.36	0.7223	-0.006	0.004
Male vs. Female	-0.029	0.018	-1.57	0.1163	-0.064	0.007

significant difference regarding accuracy.

Additionally, data were divided not only by levels but also by lives[16]. Therefore, an analysis to find the correlation between the number of lives (per subject and level) and Condition, Level, Age, and Sex was calculated. The analysis used a mixed-effects cumulative logit model. These logistic random effects models are a popular tool to analyze multilevel data, also called hierarchical data with a binary or ordinal outcome. Ordinal data are a categorical variable in which levels have a natural ordering (e.g., light, medium, heavy) (Liu and Hedeker, 2006). Furthermore, data can have a hierarchical or clustered structure (e.g., school, families) or repeatedly measured across time. In consequence, using mixed-effects regression models, are helpful to correlate all the variables. In particular, a logistic-based model is helpful for the analysis of ordinal data and multileveled (Liu and Hedeker, 2006). Therefore, in the analysis, the target variable was the number of lives (per participant and level), and covariates condition (auditory or tactile), level (Level 1 and Level 2), age (twenty to thirty-one years old), and sex (female, male). Participants' random effect was incorporated because the taps (absolute

[16]Lives are considered as each life of the three lives that each player has to finish a level.

accuracy) from the same participant may be correlated. The results of the model are shown in Table 3.7.

TABLE 3.7: *Results from the mixed effects cumulative logistic model, bold font highlights significant effects.*

Effect	Estimate	Standard Error	t Value	Pr >|t|	Lower 95% CL	Upper 95% CL
Intercept - 2 lives	3.9929	3.0562	1.31	0.2043	-2.3294	10.3152
Intercept - 3 lives	4.9726	3.0903	1.61	0.1212	-1.4202	11.3654
Tactile vs. Auditory	1.1565	0.6982	1.66	0.1101	-0.2814	2.5944
Level 2 vs. Level 1	**2.8720**	**0.7963**	**3.61**	**0.0014**	**1.2319**	**4.5121**
Age	-0.1781	0.1198	-1.49	0.1496	-0.4249	0.0686
Male vs. Female	-0.6793	0.8214	-0.83	0.4161	-2.3710	1.0124

Overall results (Table 3.7) indicate a significant association between level and number of lives (p-value = 0.0014), where players needed more lives to finish Level 2 than to finish Level 1. However, the association between stimuli and numbers of lives was not significant.

The result from the task load measured by NASA TLX was not statistically significant. The mean for the auditory condition was 44.9, with a standard deviation of 14.30, while the mean for the tactile condition was 46.41 with a standard deviation of 17.74 (Table 3.8). Regarding the user experience questionnaire (UMUX), both conditions were positioned in the upper quarter of scores; auditory condition obtained a mean of 3.85, and tactile condition reached a mean of 4.00 from the five-point scale. Concerning the motivation of players (IMI), participants rated different subscales with similar points between auditory and tactile conditions (Table 3.8). In the IMI questionnaire, enjoyment and effort were rated higher than competence and tension. Nevertheless, neither results from the UMUX questionnaire, nor results from the IMI questionnaire were statistically significant. Table 3.8 displays overall results from questionnaires.

TABLE 3.8: *Questionnaire results of NASA TLX ([0;100], the lower the better), UMUX Lite and IMI (each [1;5], the higher the better) (mean ± standard deviation (SD)).*

	Tactile	**Auditory**
TLX	46.41 ± 17.74	44.94 ± 14.30
UMUX-Lite	3.85 ± 0.66	4.00 ± 0.85
IMI-Enjoyment	3.29 ± 0.71	3.40 ± 0.44
IMI-Competence	2.26 ± 0.80	2.44 ± 0.87
IMI-Effort	3.77 ± 0.74	3.92 ± 0.56
IMI-Tension	2.96 ± 0.77	2.91 ± 1.02

Pre-Test and Post-Test. For pre-test and post-test, mean and standard deviation were calculated considering absolute values. In the pre-test, for the auditory condition, a mean of 0.46s was obtained with a standard deviation of 0.15s. Whereas for the tactile condition, the mean was of 0.41s with a standard deviation of 0.14s. For the post-tests auditory condition, the mean was of 0.44s, with a standard deviation of 0.16s. Whereas for the tactile condition, a mean of 0.33s was obtained with a standard deviation of 0.12s. In addition, a boxplot was generated to illustrate the dispersion of data (Fig.3.19).

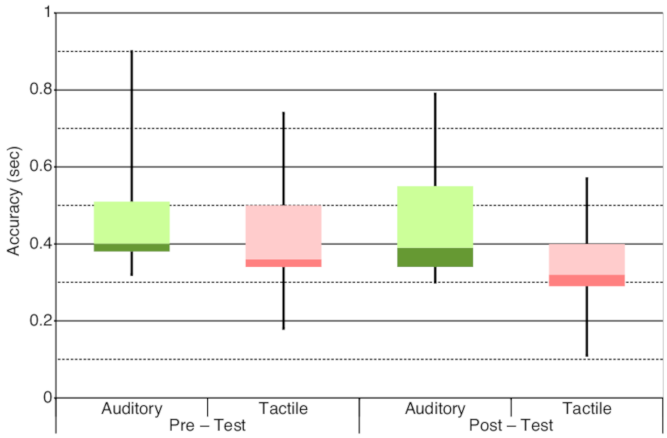


FIGURE 3.19: *Boxplot of the auditory and tactile conditions in Level 1 and Level 2. The boxplot shows accuracy of participants and the dispersion of data (absolute values), based on the minimum and maximum values, median, first and third quartile, and possible outliers.*

Participants with tactile conditions were more accurate in the pre-test (0.36s) and post-test (0.32s) than participants with the auditory condition. The pre-test was of 0.40s, and the post-test was of 0.39s. Furthermore, an improvement in participants' accuracy can be distinguished in the post-test for both conditions. In this regard, the auditory condition's accuracy was of 0.39s in the post-test and 0.40s in the pre-test. For the tactile condition, accuracy was of 0.32s in the post-test and 0.36s in the pre-test.

A linear mixed-effect model calculated the association between absolute accuracy (target variable), and stimuli, time (pre/post-test), age, and gender (covariates). Participants were the random effect because data from the same player may be correlated. Table 3.9 presents results of the model. Possible associations between accuracy and time (in pre-test or post-test) or accuracy and stimulus are not statistically significant. However, mean accuracy was estimated to be better in the tactile condition than the auditory condition. The participants were more accurate in the post-test than the pre-test (adjusted for all model variables). Moreover, the stimulus-time-interaction parameter was estimated to be -0.063s, indicating an improvement in accuracy from the pre-test to post-test with the tactile group, but this association was not statistically significant.

TABLE 3.9: *Results of the Pre and Post Test from the linear mixed regression for accuracy.*

Effect	**Estimate**	**Standard Error**	**t Value**	**Pr >|t|**	**Lower 95% CL**	**Upper 95% CL**
Intercept	0.276	0.222	1.24	0.2268	-0.184	0.735
Tactile vs. Auditory	-0.045	0.054	-0.83	0.405	-0.151	0.061
Post vs. Pre Test	-0.022	0.027	-0.81	0.4207	-0.075	0.031
Male vs. Female	0.000	0.058	0	0.9972	-0.113	0.113
Age	0.007	0.009	0.86	0.3919	-0.010	0.024
Stimuli - Time - Interaction	-0.063	0.039	-1.61	0.1071	-0.139	0.014

3.4.3 Discussion

As mentioned in the literature review, humans tend to better synchronize with auditory stimuli than with any other stimuli (Patel et al., 2005; Chan and Ng, 2012). However, the results of Huang et al. (2013), and Brochard et al. (2008) found that humans perceive music meter via tactile stimuli in a similar manner to that perceived via auditory stimuli. The first hypothesis in this study sought to determine whether players' performance using tactile stimuli is similar to the performance of players using auditory stimuli (H4). The current study found that participants with the tactile condition had an absolute deviation from the perfect accuracy (0s) 0.037s (p-value=0.0177) which was higher than participants with the auditory condition (Table 3.6). Therefore, participants with the auditory condition were more accurate, which confirms hypothesis H4. Several reports have shown that this accuracy value is inside the range that the literature reports (0.030s to 0.068s) (Diederich and Colonius, 2004; Hanson et al., 2009).

What is surprising is that an association between the accuracy and the number of patterns performed by participants was found. Each time that participant moves forward by pressing the space bar, the participant had a delay. This delay is 0.001s (p-value=0.0007) per tap. This association might result from a fatigue effect during their performance (Table 3.6).

The results of this study did not show any statistical significance regarding the accuracy of participants and the presence or absence of supportive game-elements. Therefore, H5 (*The absence of game-elements hinders the performance of the players in both conditions*) was not confirmed. Nevertheless, there is a significant association between levels and the number of lives (p-value=0.0014) (Table 3.7). The number of lives used in Level 2 was significantly higher than in Level 1. The number of participants that used three lives to achieve Level 2 was 85.7% for the auditory condition and 100% for the tactile condition (Fig.3.17). In contrast, in Level 1 (key game-elements uncovered), 28% and 53.85% of participants used three lives for the auditory and tactile conditions, respectively (Fig.3.17). This relationship may partly explain that participants in Level

1 most likely performed according to platforms and gaps, and stimuli faded into the background. Furthermore, independently of the level, players who used two or three lives improved their accuracy in life 2 (0.028s, p-value=0.0001) and life 3 (0,032s, p-value=0.0002) in comparison to life 1 (Table 3.6). Nevertheless, it was not the case between life two and life 3.

Contrary to expectations, this study did not find a statistically significant user experience, either with the auditory or tactile conditions. Therefore, hypothesis H6 (*The user experience is higher in participants using the auditory stimuli*) was not confirmed. However, some insight from the data showed that either auditory or tactile conditions are appropriate for users (Table 3.8). Moreover, participants considered that their task load, usability, and enjoyment were in balance. Thus, neither tactile stimuli nor auditory stimuli made players feel less competent during gameplay. However, without significant differences, caution must be applied to these assumptions.

An initial objective of the study was to identify an improvement in participants' rhythm skills after playing the game. Tests were designed to include beat tapping and sequences processing. In the beat tapping section, participants were required to tap in synchrony with a rhythm pattern. Whereas in the sequences processing section, participants were required to reproduce a rhythm sequence without any stimuli. Nevertheless, it seems that instructions were not clear enough for participants; consequently, data obtained from the sequences processing section were flawed. In order to recover some data samples, the analysis considered only data from the tapping section. The results of the pre-test and post-test indicate that no statistical significance was found between them. However, observations of the results showed that participants improved their accuracy by 0.063s during the post-test compared to the pre-test (Table 3.9). Furthermore, better accuracy of participants with the tactile condition than with the auditory condition was observed. These findings were also reported by Seim et al. (2016) and Bégel et al. (2017). These data must be interpreted with caution because it was not statistically significant. However, a possible explanation for these results may be the high values for auditory and tactile conditions (mean=0.41s and mean=0.44s, respectively), which were almost three times greater than those reported in the literature (Bangert et al., 2001; Diederich and Colonius, 2004; Hanson et al., 2009; Huang et al., 2013). These values were reasonable in the analysis because to grade equal performance of both conditions, a penalty of 0.9s was added for missing or extra taps. Therefore, these values can mask the original reaction times.

While preliminary, these findings suggest that the use of tactile stimuli is adequate in activities such as following rhythm patterns in a video game environment. Video games can incorporate vibrotactile stimuli to give a more immersive experience in the game as the main element in keeping players' attention. As the results show, players' performance depends not only on the stimuli but also on the game-elements contained in a rhythm game. Therefore, it is mandatory to find a balance between game-elements and stimuli to keep players' attention in rhythm sequences. The findings presented in this case study also suggest incorporating vibrotactile stimuli in designing game technology for people with special needs, e.g., hearing impair individuals. Consequently, a fundamental activity linked to reading, speaking, sporting, and music, such as rhythm skills, can be acquired by everyone through training.

Further research should be undertaken to investigate the effects of rhythm learning in a long-term study. Furthermore, it could be interesting to present the prototype in deaf communities to gather their opinions regarding using this technology to improve their music skills. A comparative regarding the performance of people with and without hearing diseases could also help to understand, from a behavioral psychology viewpoint, how people with a hearing disease fill the gap of a missing sense in activities such as rhythm acquisition.

3.5 Chapter Summary

This chapter described the use of rhythm games not only as an entertainment tool, but also as a tool to improve rhythm skills in players. This chapter addressed two case studies in which the behavior of players was analyzed. The case studies used the game *Jump'n'Rhythm* in two versions: 1) using auditory and visual stimuli and 2) using auditory and tactile stimuli. Results indicated that auditory stimuli are more suitable in rhythm synchronization tasks; however, synchronization with tactile stimuli is also feasible. Moreover, game-elements play a vital role in the performance of players. This chapter, together with chapter five, opens a panorama to study music video games from a multidisciplinary perspective, involving areas such as psychology, computer science, and pedagogy.

Chapter No. 4

Pitch: the Second Step of Music Learning

This chapter presents a game prototype for pitch recognition based on auditory-motor theory. The first section talks about the use of video games for music learning, particularly for pitch recognition. Afterwards, it describes some findings regarding the benefit of using music video games and points out some aspects to consider for designing games for training players in pitch recognition. To understand and contribute to filling the gap, this chapter deals with the development process of a prototype for pitch recognition. The prototype aims to analyze the learning effect through game experience and to identify an association between body movement and stimuli of the video game. This chapter finishes with a discussion and conclusion of the general use of music video games for pitch recognition.

4.1 Auditory-Motor Interaction in a Video Game

Games have been shown to be motivational. They support the players' body's learning and training to acquire specific skills (Malaka, 2014). Investigations propose that music video games such as *Guitar Hero* (Harmonix et al., 2005) can help students improve their musical skills and support the players' embodied competence (de Castell et al., 2014). An important aspect to consider in the design of music video games should be the embody cognition, action and reaction during the music video game. It should balance learning and gameplay principles to design video games for learning, where the integration of the embodied cognition process should be paramount to achieving the learning goals. However, how can we integrate these mechanical elements of gameplay with the human body's natural behavior in a music context? Following this question, a video game to train pitch recognition was designed and developed, where the main element is the auditory-motor interaction. This game was analyzed using the Learning Mechanics - Game Mechanics Model (LM-GM) (Arnab et al., 2015) to determine whether the prototype accomplishes the learning goals for pitch recognition. Afterward, a case study was conducted to evaluate how gamification affects the subjective experience of players with the following hypothesis:

- H7: Gamification of a pitch recognition task is more attractive to youths than a non-gamified task.
- H8: Users learn a task such as pitch recognition by the use of a video game.

4.2 Design of a Game for Pitch Recognition

To reach the hypothesis pointed out, a game was designed by Jöllenbeck and Alvarez-Molina et al. (2018), where player experiences and their relation to learning goals are the central part of the design. In addition, the game considers the pitch recognition of players as one of the learning goals.

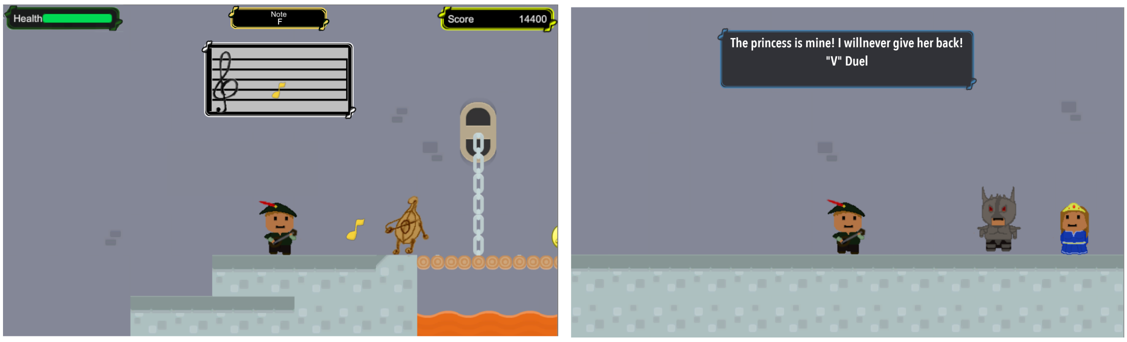
To achieve the learning goals pointed out, it is necessary to consider a methodology centered on the design of video games-based learning. Moreover, for the design of educational video games, it is important to work with a multidisciplinary group, where communication is essential. Therefore, a methodology that all team members can understand is a fundamental aspect of designing educational games. Thus, the methodology for developing educational games based on interactive screenplays proposed by Prieto et al. (2015) is based on the narrative of video games and follows a series of items organized into chapters and scenes. Other elements are progressively added to the script (e.g., scenarios, characters, challenges). The proposal of Prieto et al. (2015) encompasses everyone involved in the development of a video game. Thereby, it is understandable for everyone, including technicians, designers, educators. Furthermore, narrative as the core helps people from all fields create the dynamics of the game and is complemented by the game designers. The methodology follows a series of items ordered from the pre-design to the scene design elements.

Therefore, the prototype presented in this chapter follows Prieto et al. (2015) methodology considering the order proposed, which is presented by items (Alvarez-Molina et al., 2018):

1. Educational objectives: The specific educational aim is pitch recognition; therefore, it is based on the auditory-motor interaction.

2. Type of video game:The game is designed for people interested in learning music. The game genre is a 2D scrolling platform with a basic narrative (Fig.4.1), where the player interacts with game-mechanics through the avatar. This interaction is via a computer keyboard in order to assign to each sound its corresponding key based on the auditory-motor theory. The game incorporates only pitches from C to D in octave 5 with a piano timbre.

3. The story and main characters: Story is about a kingdom that is cursed by a wizard, and all residents become musical instruments. Only one resident, the bard (avatar), is not cursed, and his mission is to rescue the princess and eliminate wiz-

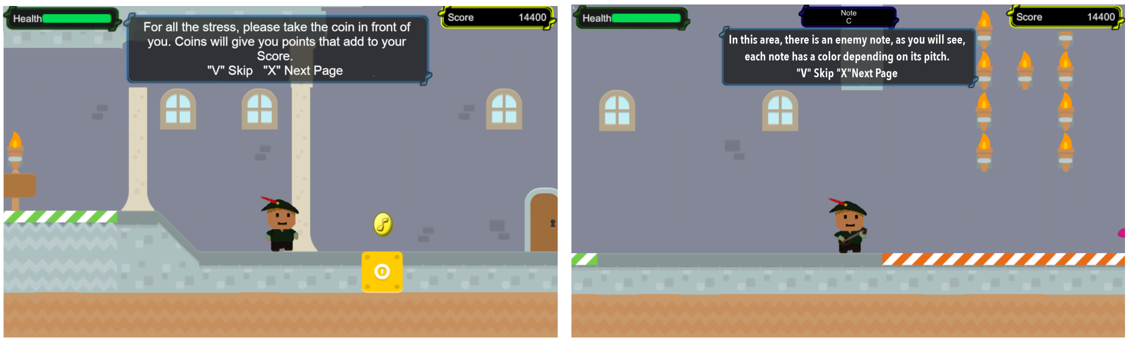


(A) *The avatar confronting a minion.* (B) *The wizard, the bard and the princess*

FIGURE 4.1: *Two scenes of the game to illustrate the story and some of the characters of the game.* ***A)*** *The avatar confronting a minion. The minion launches a music note to attack the avatar, inside a castle with lava.* ***B)*** *The wizard interacts with the avatar by a dialog box.*

ard (boss) to break the curse. Thus, the avatar needs to eliminate the opponent (minion) to get access to the boss (Fig.4.1).

4. Chapter design: To date, the game has only one chapter and two tutorials (Fig.4.2). The aim of the first tutorial is that players become familiar with avatar movements and game elements. The following tutorial is designed to learn how to



(A) *First tutorial* (B) *Second tutorial*

FIGURE 4.2: *In both figures, the dialog boxes guide the player for future actions.*
A) *Corresponds to the first tutorial where the player practices how to move the avatar, the dialog indicates the purpose of the coins during the game.* ***B)*** *Second tutorial, where the player is familiarized with the instrument and how to eliminate the opponent. For instance, dialog box explains to the player why musical notes appear, and why they have different colors.*

FIGURE 4.3: *The figure displays a forest scene where the avatar faces a minion.* ***A)*** *Indicates the health bar; every time that a music note touches the avatar the value decreases.* ***B)*** *A box shows note name (B) launched by the minion. The edge color depends on the note (Appendix B).* ***C)*** *Score Bar that indicates the points collected from the coins. D) The arrows points to music notes that the minion launched and its corresponding position on the stave; the note is colored according to its tone (Appendix B).*

throw pitches in order to eliminate a minion.

5. Scenario: There are two scrolling sceneries (castle and forest) and a static scenario (with the boss).

6. Characters: The avatar moves by pressing the arrow keys and space bar, as well as uses the instrument by selecting the corresponding pitch (key-pitch = 1-C, 2-D, 3-E, 4-F, 5-G, 6-A, 7-B). The minion moves forward and throws a pitch. The boss is a static character who only plays simple melodies.

7. Play challenges: The player must overcome a duel, an avatar vs. a minion (Fig.4.3). The challenge here is to match the pitch of the minion with the pitch of the avatar. Pitch generated by the minion is random; therefore, the player does not know which pitch to match next. Moreover, the avatar is required to collect coins and recover its health by picking up health-boxes. Finally, in the boss duel, the player

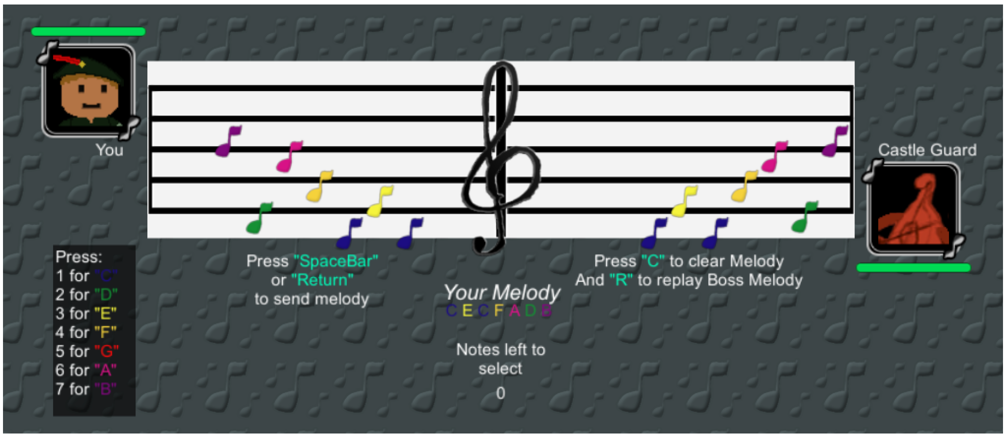


FIGURE 4.4: *The duel screen consists of some elements: a health bar and the face of the avatar and a health bar and the face of the boss. Every time either the avatar or the boss loses a game melody sequence, the bar decreases depending on who loses. In the left-bottom, a list is displayed with each note name and its corresponding key on the keyboard. Moreover, the note name has the same color as its corresponding tone. The boss throws its melody sequence from right to left, whereas the avatar launches its melody sequence from left to right. The screen displays instructions on how to perform this part of the game.*

should identify a sequence of pitches played randomly by the boss so that the player must reproduce the same pitch sequence in order to beat the boss (Fig.4.4).

8. Educational challenges and assessment: For pitch recognition, a confrontation between avatar and minion was designed. The player has an auditory stimulus from minion and generates a motor reaction. This design is based on auditory-motor theory (Bangert et al., 2001; Lahav et al., 2005, 2007) and drill-and-practice strategy, which promotes the acquisition of knowledge or skills through repetitive practice. As support for beginners, each pitch has a color; thereby, players can identify pitch with color as a starting point.

9. Emotions: The aim of the game design is to obtain significant results regarding game experience and learning. In particular, the game seeks a balance between challenge, flow, tension, negative affect, and other factors that affect learning. To achieve this, the game considers different methods of feedback, narrative, challenges, and flow.

10. Adaptation design: The game can be adapted to different platforms such as tablets, smartphones, and computers. However, it is important to keep the association between the pitch and its corresponding finger.

11. Collaboration design: The design of the game is for only individual use. The option of a collaborative or multi-player game is not available.

During the design process, a canvas based on Game Design Canvas proposed by

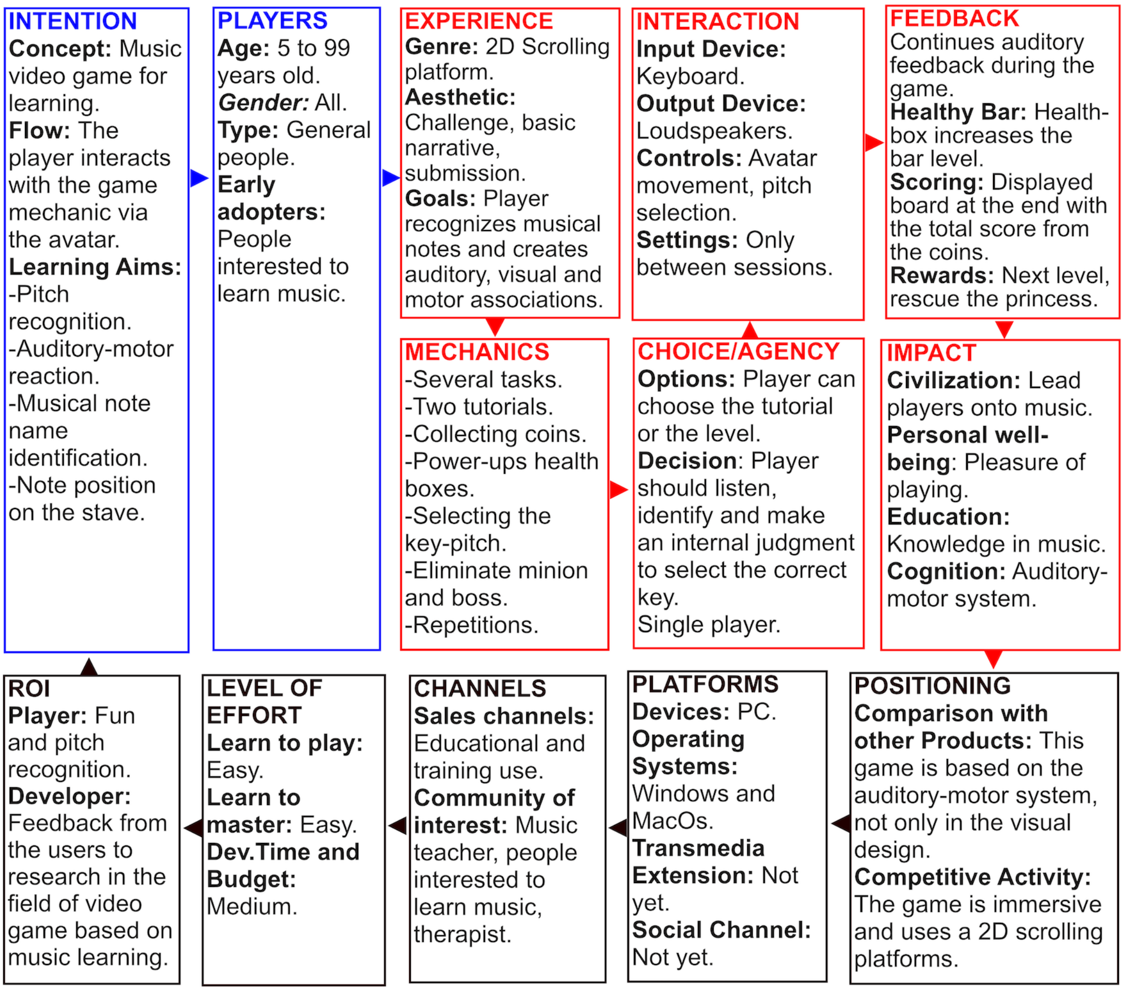


FIGURE 4.5: *Game design canvas using the canvas design proposed by Carey (2015) with the game elements presented in this work. Blue edges indicate the setup part, red edges point out the design part and black edges indicate the consideration part (Alvarez-Molina et al., 2018).*

Carey (2015) captures design thinking and represents the design in a simple way. This representation forces designers to capture game design in a few words and to focus on one part of the game at a time. The idea is centered on users because it defines the player as a first and primary step. Therefore, a simple representation of design can help designers create possible interactions and perfect design. Canvas considers three central parts: setup, design, and considerations. *Setup part* (blue edges in Fig. 4.5) is the intention of the game and the type of players who are going to play the game. For instance, the game presented in this chapter is designed to support music learning, in particular, pitch recognition based on muscle memory. The *design part* (red edges in Fig. 4.5) encompasses different items such as experience, interaction, feedback, mechanics of game, choices, and the impact of creating the experience. On the other hand, the *considerations part* (black edges in Fig. 4.5) shows which other factors affect the design, marketing, and development (Fig. 4.5).

4.3 The Use of LM-GM Model to Evaluate the Game Prototype

To analyze the content of the learning components of the game it was necessary to use a model based on the connection between pedagogical goals and game elements of the video game. LM-GM model (Arnab et al., 2015) allows the user to find the relation between learning-mechanics (LM) and game-mechanics (GM). The dynamic way of the LM-GM model supports designers in identifying learning-mechanics during each scene and each action of the game. Thus, a map represents learning-mechanics and its relation with equivalent game-mechanics during the flow of actions.

Gender and simple narrative of the game allow representing the flow of actions of video games in a simple way by the LM-GM map (Fig. 4.6-A). Pitch recognition training is based on the association between pitch and its corresponding finger (auditory-motor system). Therefore, the recognition will generate by repetitive practice. The LM-GM map, therefore, is centered on the action of listening to pitch, identifying, and choosing the correct key, and pressing it. The second step is to determine which one of game-

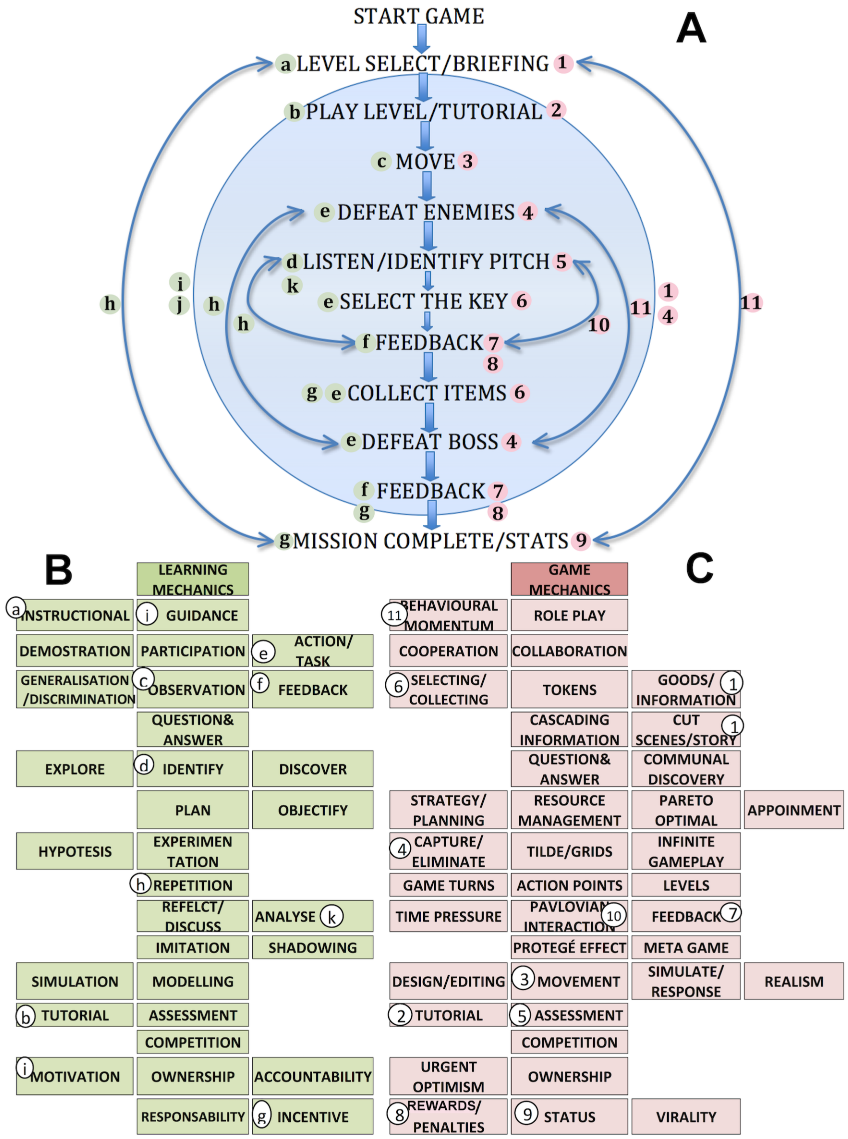


FIGURE 4.6: ***A)*** *Representation of the dynamic view of the LM-GM model adapted from Arnab et al. (2015) to the game presented in this work.* ***B)*** *Learning-Mechanics node shows a list of learning elements extracted by Arnab et al. (2015) mainly from educational theories.* ***C)*** *A list of Game-Mechanics elements extracted by Arnab et al. (2015) from game theories is displayed.*

mechanics belongs to or is related to learning-mechanics. To find this relation, the LM-GM model proposes two tables (Fig. 4.6): a table that contains learning-mechanics (LM) (Fig. 4.6-B) with central aspects of learning processes according to pedagogical principles and a table that covers main game-mechanics (GM) (Fig. 4.6-C) used in most of the game designs.

For instance, the first main action in the game is *level select/briefing*, that a menu provides options such as two tutorials and start the game. Thus, in the map (Fig. 4.6), this action appears first, and contains *instructional* learning-mechanic because it gives a general rule about how the player should interact with the game through dialog windows. The next section is a *tutorial*, which appears in both tables. In this tutorial, the player can practice and learn interaction with the avatar and environment. The move is considered a game-mechanic; the avatar can move through different scenes. The relation of movement with learning-mechanics is the observation. The player can observe the environment and analyze various alternatives to finish the level; moreover, the movement gives user dynamic game experience. The central part of action and learning in the game is provided by *defeating enemies*, where the task (LM) is to eliminate (GM) enemies. To achieve this, the player enters a cycle of *listening/identify pitch*, select key, and receive auditory and visual feedback. For the action listen/identify (LM), the enemy launches a musical note, and the player must to listen to the note; then, for the launched note, the player needs to identify the note. The player then presses the corresponding key on the keyboard according to the note that the player considers correct. The design of this cycle of actions and tasks considers the auditory-motor system in a similar way to Lahav et al. (2005) study, where participants associate a specific finger to a specific pitch to learn a small piano melody. Furthermore, on one hand, feedback (GM) for the player appears each time the player fails, reducing the health bar (GM). On the other hand, when the player successfully find the correct key, the points increase (LM-GM), and the enemy disappears. The cycle is repeated several times during the gameplay. Therefore, the repetitions are the learning-mechanic, and the Pavlovian interaction is the game-mechanic.

During the gameplay, the avatar needs to collect (GM) items, which increases points to the score. These items are an incentive (LM) for the player. To achieve a level, the player must accomplish an additional task, a duel, which consists of reproducing the same melody that the boss generates in order to eliminate the boss. The duel, therefore, considers game-mechanics of eliminating and its relationship with learning-mechanics of action/task. Feedback considers game-mechanics of rewards or penalties via health bars of the avatar and the boss. Actions on the map finish with *mission complete/status*, which gives the player a final score and congratulations on rescuing the process. Then, the cycle starts again. As a result of using the LM-GM model to analyze the pitch recognition prototype, some points of conversion were found between learning-mechanics and game-mechanics. Moreover, the model allows us to identify the main action, how this action is developed, and whether it can support pitch recognition.

4.3.1 Case Study

The present case study explores the participants' experience using either the pitch recognition prototype or a web application to train pitch recognition. The study is based on the seven subscales of the Game Experience Questionnaire (IJsselsteijn et al., 2008) and the participants' comments and reactions. Therefore the group of participants was divided. One group played the pitch game, and the other group used the web application. The study attempted to validate the hypotheses pointed out in the section above. Moreover, the comments of participants will give guidelines for improving the game and for future designs.

Participants. The study was conducted in a middle school in Coatepec City, Mexico. Because all participants were minors, the Principal of the school informed the parents of the study, which was considered an extracurricular for the students. Participation was not mandatory; neither did it affect any grades for the participants. Sixteen participants enrolled in the middle school (included students between 12 to 15 years of age) participated in this study. The sample contained 4 males and 12 females. They were divided into two groups, where each group had two males and six women. One group used the pitch recognition prototype (game), and the other group used a web application to train pitch recognition. They reported neither auditory nor physical disabilities.

Additionally, the four males and five females played video games. Ten participants of the entire group have had contact with a musical instrument. Three of them have had contact for less than three years, and three participants have had contact for more than four years; however, only two could read music scores.

Apparatus. The game was developed using the Unity[1] engine and was played on a Mac Book Pro with an Intel Core i5 processor, 8Gb of memory, and a NVIDIA GeForce graphics board. The screen used to visualize the game was a 19 inch Acer V193W monitor with a resolution of 1440 $\times$ 900 px. The sound was transmitted to the participant by external computer desktop speakers Genius 2.0 with a frequency response of 100 - 20KHz. For performing the game, an external mechanical keyboard and an optical mouse connected via USB to the computer was utilized.

The web application used for the study was developed by musictheory.net (Musictheory.net, 2000) to train pitch recognition. The application uses piano sounds; moreover, it could be configured to choose the scale and number of pitches for training. The interaction with the application is by a computer mouse. The application emitted a pitch. In a panel (Fig. 4.7), the player chooses the note that he or she considers to be the sounding pitch. The user can hear the question note each time that he or she presses the question note box. Moreover, in the panel, there is a reference note box (for this study a C). Thus, the person can compare the question pitch with the reference note to support (by the construction of an interval of the two pitches) the identification of the question pitch. Red color gives feedback, whether the chosen note was incorrect, the box becomes red. Once the user selects the correct answer, the next question pitch will sound. A score indicates how often the user chooses correct notes vs. incorrect notes. These two values will give the percentage of accuracy.

Procedure. The experiment was conducted at the Simon Bolivar middle school in Mexico in a private room with good lighting. The participation of each student consisted of a single individual session. In this session, the participants remained seated

[1]Unity Technologies. (2005). *Unity*. [Game Engine]. Available in `https://unity.com/`

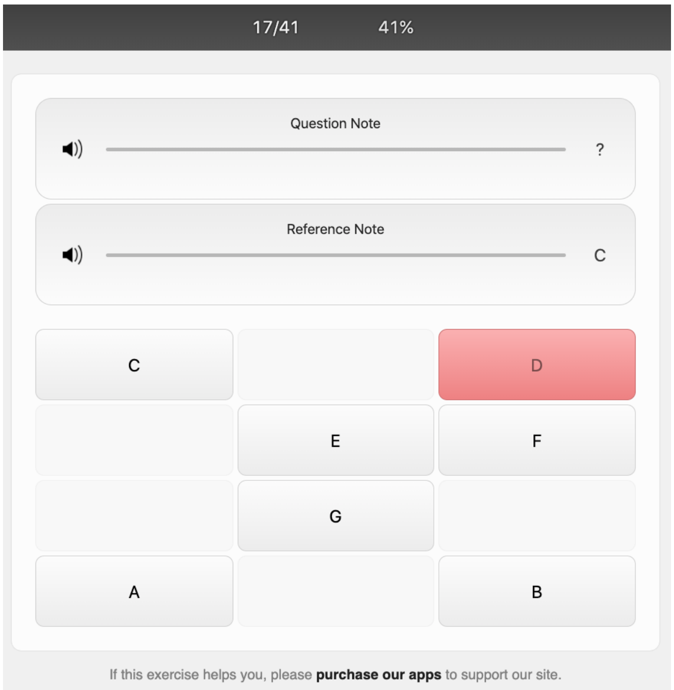
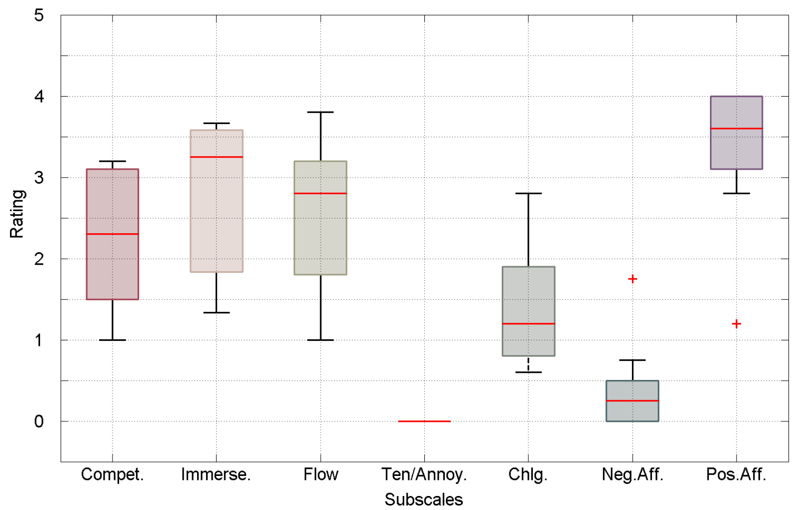
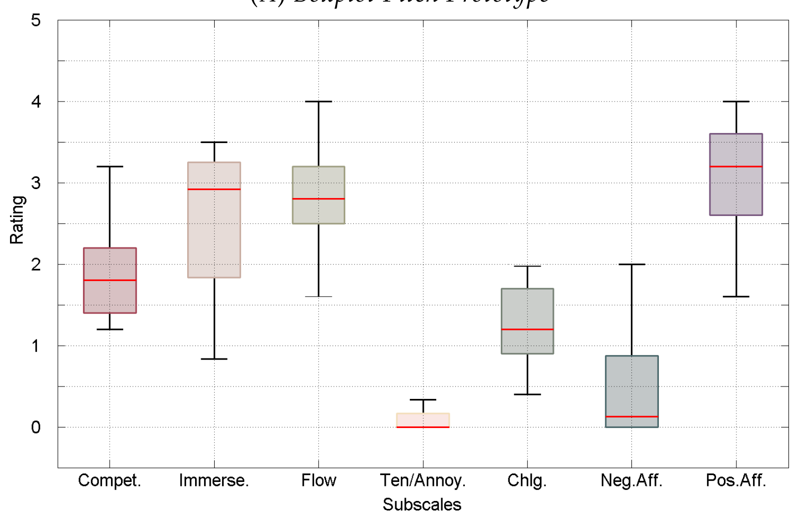


FIGURE 4.7: *In-game screen of the web application musictheory.net (Musictheory.net, 2000) where the player listens to the pitch given by the game and the player needs to choose the correct name of the note that corresponds with the listening tone. The figure shows the D box in red, indicating that is not the correct answer.*

in front of a monitor and computer keyboard. The experiment was a between-subjects design. Participants performed only one of the two conditions (pitch game or web application). After a brief introduction about the study, participants completed a profile questionnaire. Subsequently, participants played either the pitch game or the web application for 20 minutes. When gameplay was over, participants completed a Game Experience Questionnaire (GEQ) (IJsselsteijn et al., 2008). This questionnaire measures player experience in terms of subscales *Competence*, *Immersion*, *Flow*, *Tension/Annoyance*, *Challenge*, *Negative Affect*, and *Positive Affect* related to learning (Poels et al., 2007; Hamari et al., 2016; Schiefele, 2001; Fullagar et al., 2013). The core module contains thirty-three simple questions covering seven subscales. The answer is given on a five-point intensity scale ranging from zero (not agreeing with statement) to four (completely agreeing with statement). The GEQ was used because of its wide use in previous studies with different game genres and its multidimensional structure. In the last part of the evaluation, participants gave their impressions of the game.



(A) *Boxplot Pitch Prototype*



(B) *Boxplot Web Application*

FIGURE 4.8: *The boxplot shows the dispersion of the absolute value of data from the participants, based on the minimum and maximum values, median, first and third quartile, and possible outliers. Note the following abbreviations: Compet (Competence); Immers (Immersion); Ten (Tension); Annoy (Annoyance); Chlg (Challenge); Neg.Aff (Negative Affect); Pos.Aff (Positive Affect).*

4.3.2 Results

Results obtained from GEQ were processed using Matlab. The data distribution of the different subscales between the pitch game and web application was compared using a boxplot (Fig. 4.8). Boxplot helps determine where the distributions lie concerning one another by plotting the quantiles side by side. Therefore, for the pitch game, the boxplot shows a data distribution with a median of 2.3 for the subscale *Competence*; in contrast, for the web application, the median was 1.8. The subscale *Immersion* had a median of 3.25 for the pitch game; in contrast, the median for the web application was 2.91. Moreover, for the pitch game and web application, the median for subscale *Flow* was 2.8. Likewise, the subscale *Tension/Annoying* had the same median of 0 for the pitch game and web application. Meanwhile, a median of 1.2 for the subscale *Challenge* was found for either pitch game and the web application. For the subscale *Negative Affect*, the median was 0.25 for pitch game and a median of 0.125 for the web application. In contrast, the subscale *Positive Affect* had a median of 3.6 for pitch game and a median of 3.2 for the web application. Furthermore, the pitch game's boxplot displays two outliers, one in the subscale *Negative Affect* (1.75) and the other in the subscale *Positive Affect* (1.2). For the subsequent analysis, the outliers were removed.

Afterwards, the mean and standard deviation were calculated (Fig. 4.9). As a result, subscale *Competence* showed a higher (2.25±0.85) rate in the pitch game than in the web application (1.9±0.65). Similarly, *Immersion* was higher (2.79±1.00) in the pitch game than in the web application (2.54±1.03). The subscale *Challenge* showed a higher rate for the pitch game (1.4±0.76) than for the web application (1.25±0.55). Furthermore, the pitch game was higher (3.63±0.43) than the web application (3.05±0.80) for the subscale *Positive Affect*. In contrast, subscale *Flow* was lower for the pitch game (2.55±0.96) than the web application (2.83±0.69). Similarly, the pitch game (0.13±0.14) had a lower *Negative Affect* rate than the web application (0.50±0.71). Additionally, the subscale *Tension/Annoyance* had a lower rate for the pitch game (0.00±0.00) than for the web application (0.08±0.15).

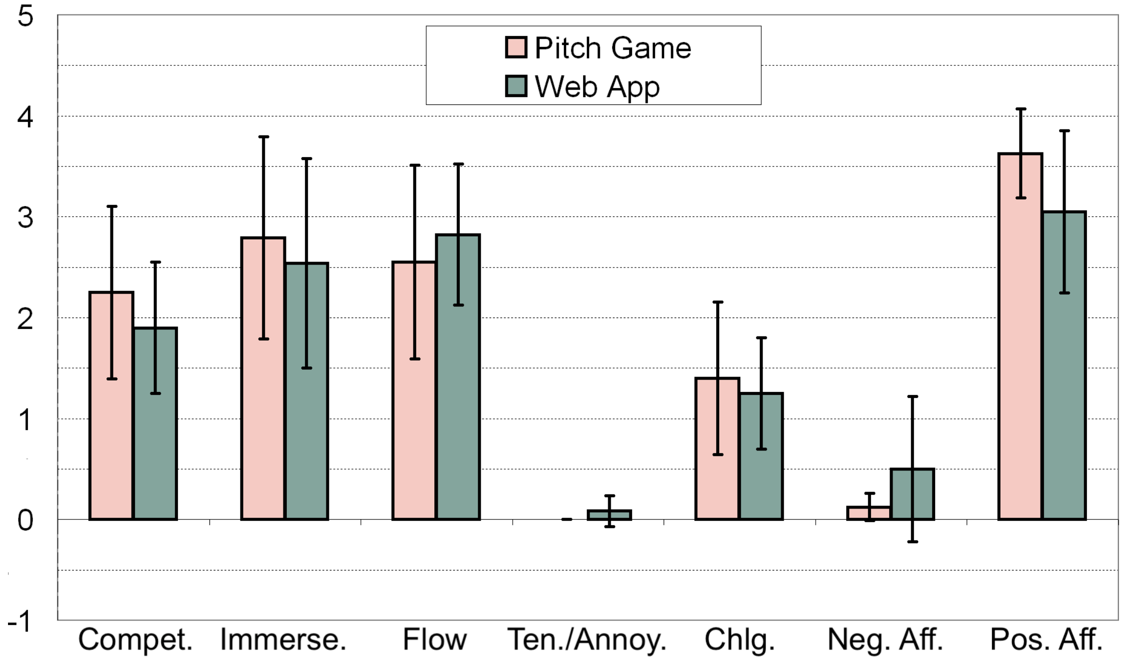


FIGURE 4.9: *Results of the GEQ. The graphic shows the mean with the standard deviation of each item of the GEQ. Note the following abbreviations: Compet (Competence); Immers (Immersion); Ten (Tension); Annoy (Annoyance); Chlg (Challenge); Neg.Aff (Negative Affect); Pos.Aff (Positive Affect).*

Subsequently, a two sample t-test was conducted between the pitch game and the web application. Contrary to expectations, this study did not find a significant difference between the subscales of the pitch game and subscales of the web application. The t-test found no significant differences in mean scores in the pitch game and web application for subscales: *Competence* [t(14)=0.92, p-value=0.37]; *Immersion* [t(14)=0.49, p-value=0.63]; *Flow* [t(14)=-0.65, p-value=0.52]; *Tension/Annoyed* [t(14)=-1.53, p-value=0.15]; *Challenge* [t(14)=0.45, p-value=0.66]; *Negative Affect* [t(12)=-1.25, p-value=0.24]; *Positive Affect* [t(12)=1.69, p-value=0.11].

A combination of quantitative and qualitative approaches was used in the data analysis. Therefore, for qualitative analysis, screen records and comments of each player were evaluated. In these records, it is worth noticing that during the gameplay using the pitch game, participants emitted short verbal expressions each time that something happened to the avatar. Whereas, participants using the web application performed the task in almost complete silence, without verbal expressions. After the gameplay, some participants expressed their comments. For instance, on one hand, participants using the pitch game commented:

- After 20 min of play, *"I want to play more."* (VG04)
- *"I think I can learn pitch."* (VG02)
- *"I like the game."* (VG07)
- *"It is a long time ago that I didn't play this kind of video game. I usually play games with more complex stories and worlds. I didn't remember that it feels so well to play these games with a simple story. It was a curious and different experience. I would like to know if there will be a version of this game on the market."* (VG08)

On the other hand, participants using the web application commented:

- *"Where can I find this application?"*(WA03)
- *"It was boring"* (WA06)
- In minute 14, *"I'm tired, how many times will the task be repeated?'*(WA07)

No further comments were expressed. However, during the experiment the tester made some observations. For instance, after some minutes of starting the gameplay, participants using the pitch game started to associate low pitches with the left position keys and the high pitches with the right position keys. Participants interacted more with the tester during the pitch game session than participants using the web application. Participants using the web application seemed bored after have way through the game. Three participants finished the game in less than 20 min; the remaining players kept playing till they accomplished the 20 minutes.

4.3.3 Discussion

Interactive music interfaces based on learning support people in acquiring music skills. In video games, the learning effect will depend on game-elements and how they are related with learning.

An initial objective of the project was to design a prototype for pitch recognition by the auditory-motor system. The main point of the approach consists of three main elements: listen, recognize, and react. Thus, auditory-motor theory is suitable for creating and implementing different game designs. As described earlier in this chapter, auditory-motor theory appears when a minion launches a music note. The player listens to this note (listen). The player identifies the music note (recognize) and reacts by choosing a key on the computer keyboard (react). Afterward, the creation of complementary game-elements also considers the correlation between game-mechanics and learning-mechanics. Therefore, during the design process, it was necessary to use some methodologies to maintain the correlation of game-elements. The first part of the design followed the methodology proposed by Prieto et al. (2015). This methodology describes each item that a game must contain; moreover, it is comprehensible for everyone involved in the development process. Subsequently, the canvas of Carey (2015) allows designers to arrange the content of the game systematically (Fig.4.5). Moreover, it analyzes which aspects in the game are missing, such as platforms, impact, and positioning in the market. In this way, weak elements of the game can be seen, along with possible improvements. It is possible to exchange these design elements between categories to achieve learning objectives. However, the use of canvas is for general video game designs. It cannot provide specific evidence of the learning elements that a game provides. Thereby, to analyze the learning content pointed out in the methodology proposed by Prieto et al. (2015), it was necessary to use the LM-GM model (Arnab et al., 2015). This model allows relating learning-elements with game-elements of a game in a more concrete way. Results obtained from using the LM-GM model show that many game-elements of the pitch game are associated with learning. These game-elements are mainly related to pitch recognition (Fig.4.6).

For the quantitative analysis of the case study, visualization of data through the function boxplot allows understanding the distribution of data obtained from GEQ (Fig.4.8).

The present study was designed to determine the effect of game-elements in learning (*H8: Users learn a task such as pitch recognition by the use of a video game*). This study set out to assess the player experience through two conditions: pitch recognition game and web application. For the evaluation, the GEQ was filled by each participant of the two conditions. Contrary to expectations, this study did not find a significant difference in the player experience between the pitch game and the web application. Therefore, the hypothesis *H8* was not confirmed. Nevertheless, these findings will not likely be scrutinized, but there are some immediately reliable conclusions for the effect of both conditions in the participants' learning. As mentioned in the literature review, there are aspects of learning (such as engagement, motivation, immersion, and flow) that impact participants' learning (Schiefele, 2001; Strati et al., 2011; Fullagar et al., 2013; Hamari et al., 2016; Cheng et al., 2015). Although the GEQ is not completely validated to date, it has been widely applied by game researchers and practitioners to a broad scope of games and purposes (Law et al., 2018). Furthermore, subscales of the GEQ provide information about these aspects of learning.

For Csikszentmihalyi et al. (2014) and Fullagar et al. (2013), one of the factors that involve engagement is competence, defined as the extent to which players feel strong and skillful (Gajadhar et al., 2008), but it is not the parameter that affects learning (Hamari et al., 2016). Nevertheless, in the study, the feeling of competence was higher for participants using the pitch game (2.25 ± 0.85) than the web application (1.9 ± 0.65). Although the subscale *Immersion* was higher in the pitch prototype (2.79 ± 1.0) than in the web application (2.54 ± 1.03), it is not related to learning. Instead, the feeling of immersion addresses player experience from the perspective of audiovisual execution of the game (Nacke and Lindley, 2008; Cheng et al., 2015) and does not affect the learning process directly (Hamari et al., 2016). One unanticipated finding was that a feeling of flow was higher in the web application (2.83±0.69) than in the pitch game (2.55±0.96). According to Shernoff (cited in (Hamari et al., 2016)), flow is associated with three phenomena: concentration, interest, and enjoyment; that are correlated with learning. Thus, subscale *Flow* can indicate a learning effect on participants during the performance of the presented game. These results, however, need to be interpreted with caution because they are not significant. Although *flow* was rated higher for the web application, there is a feeling of tension or annoyance for the web application (0.08±0.15) and not for the pitch game (0.0±0.0). In consequence, some feelings of frustration and irritation can be present in participants using the web application. However, frustration or irritation is less persistent and less associated with poorer learning than boredom (Baker et al., 2010). Prior studies have noted the importance of balancing challenge and skills for learning (Strati et al., 2011; Fullagar et al., 2013; Hamari et al., 2016). Therefore, for the pitch prototype, the subscale *Challenge* was higher (1.4±0.76) than for the web application (1.25±0.55).

The result in the *Positive Affect* subscale shows great fun and enjoyment of participants, either using the pitch game (3.63±0.43) or the web application (3.05±0.80). However, the subscale is associated with audiovisual and interactive design (Gajadhar et al., 2008) rather than with learning. In the case of the *Negative Affect* subscale, the pitch game shows a low rate (0.13±0.14) in comparison to the web application (0.50±0.71). This scale concerns feelings of boredom and distraction, that are indicators of poor learning (Baker et al., 2010). These results, however, need to be interpreted with caution because they are not significant.

The results suggest a slight advantage of the pitch game over the web application for pitch recognition training. Some key subscales related to learning obtained a rate according to the expecting result. Nevertheless, the results were not statistically significant. It is difficult to explain this result, but it might be related to the fact that participants were affected by the environment (school). In the GEQ, it is possible that participants felt committed to answering according to what they think was the correct answer rather than answer according to what they were feeling.

Regarding the qualitative data, the pitch game received more positive comments than the web application. Even the web application received more negative comments related to boredom. Additionally, after the gameplay, participants using the pitch game interacted (talked) more with the tester than participants performing the web application. A possible explanation for this might be that participants using the pitch game felt more confident and happier during the session. However, there is no statistically significant data to validate the hypothesis *H7* (*Gamification of a pitch recognition task is more attractive to youths than a non-gamified task*)

From the tester's observations, there is an association of low pitches with the key in the left position and the high pitches with the keys in the right position of the computer keyboard. These relationships may partly be explained by recognizing the auditory-motor system, which identified one's hand movements with a specific pitch. However, due to the indications being unclear regarding the use of a specific finger for a specific key, auditory-motor learning cannot be assessed. Additionally, using a computer keyboard for the task is probably not as good as a piano keyboard to associate a specific pitch with a specific key.

Future work must evaluate the game experience and its relation to learning as well as how participants improve their pitch recognition. In this way, two objectives are expected to be accomplished, the first one regarding the improvement of game-mechanic (GM), learning-mechanic (LM), and design more levels based on results observed in the study. The second objective is to hold a formal case study, where different variables will be measured, such as pitch recognition through a pitch-recognition-production pre-test and post-test, level of stress bio-sensors, and comparison to a control group in a traditional music lesson. Moreover, a case study is necessary to assess pitch recognition through the auditory-motor system.

In summary, the game presented in this work has the potential to be an alternative method to recognize pitch and musical notes using unconscious movements of the human body. Furthermore, evaluations that could be derived might better understand the behavior of the human body with different stimuli in a video game environment.

4.4 Chapter Summary

Player experience plays a vital role in learning. Therefore, looking to improve the learning experience in a pitch recognition task, this chapter presented a prototype to train pitch recognition. This chapter described the game-elements of the prototype and their relationship with the learning-mechanics of game. In addition, in a case study, the prototype was evaluated regarding game experience to find possible learning effects.

Chapter No. 5

Conclusion

Music is an integral part of human life, producing not only feelings and emotions, but music also influences the physiological processes. For instance, music generates body movements by action-perception processes. Therefore, music has been used in tasks where body movements are predominant such as sports, dance, and physical therapy. However, as aforementioned throughout this thesis, music is not only a physical process where different energies produce the sound that humans perceive, but music is also a process in the human mind. Without the internal mental process, music is only the physical part, without human reaction, without emotions, feelings, benefits for human life. This thesis has defined the music mind as internal mental processes that include experiences, perception, meaning, sensory, and motor systems. When the body and the music mind work together, they generate cognitive processes (e.g., learning, memory, and prediction). This interaction between music and the body is called embodied music cognition.

To date, different technological tools modify the experience of people with music. These tools allow for interaction with the music mind, stimuli, and body. However, to achieve an optimal experience, the tools must be transparent. The term transparent, in this context, refers to the technological body where the tools become part of the body. Therefore, persons do not perceive or feel the tools as external devices. Instead, they feel the tools as a part of their own bodies. The use of these technologies has allowed neuroscientists (among others) to incorporate such devices into their research. These tools support the exploration of human perception of different stimuli and the analysis of body reaction. Therefore, this thesis has considered the use of technological tools to understand human behavior with music. In particular, this thesis has addressed the use of a specific technological tool to enhance music experience -video games-. People, and in particular youths, have their first contact with music through technological tools, such as video games. Moreover, it is well-known that video games generate positive feelings and a state of immersion; consequently, the players engage with the game. Also, depending on the association of cognitive elements such as sounds, images, or body movements, players can learn from video games.

This thesis has explored the relationship between music and video games from the perspective of players' behavior.

5.1 Contribution

The Design of Music Video Games

When people think about video games, the first idea that comes to mind is that video games are comprised mainly of visual stimuli. Nevertheless, other stimuli affect players' behavior, such as auditory or tactile stimuli. For instance, visual stimulus appears in a character (Mario Bros.), and the auditory stimulus emerges when this character acts (collects a coin). A tactile stimulus reinforces the character's action (vibrates when the character performs an incorrect action). However, depending on the type of video game, some stimuli will be more suitable than others. In rhythm games, where the task synchronizes with sequences of events, auditory or tactile stimuli enhance players' performance more than visual stimuli. In games where a specific finger movement is needed for certain actions, auditory stimuli have been proven to facilitate the association between movement and action.

The first contribution of this thesis was to design and create a rhythm game (*Jump'n' Rhythm*) to study the players' reactions through different stimuli during the game. The game was designed in collaboration with Joshua Wollersheim, Dmitry Alexandrovsky, and Katya Alvarez-Molina (Alexandrovsky et al., 2016). The game-elements, game-dynamic, and game-mechanics were defined based on the prototype's purpose during the design sessions. Moreover, the project required feedback from a music expert to improve the rhythm skills of players. Therefore two music professors were contacted. Prof. Eva Verena Schmid from the University of Bremen who supported visual design, for instance, the idea of hiding some game-elements (platforms). Additionally, Prof. Uwe Seifert from the University of Cologne gave feedback on sounds (e.g., to avoid the metronome sound). Although the data acquisition was through an external method (audio recording) in the first case study, the second study's prototype obtains the data directly through the game itself. Creating an internal method for data acquisition allows researchers to use the prototype in different experiment designs. Therefore, the researcher does not need recording equipment, only needs the game.

The second contribution was designing a video game for training pitch recognition (*A Bard's Tale*). The game was designed by Aaron Jöllenbeck and Katya Alvarez-Molina (Alvarez-Molina et al., 2018). Moreover, during the design sessions, Prof. Eva Verena was consulted to obtain her feedback regarding the music elements. The prototype is based on the auditory-motor system; for each pitch, the player must to press a specific key on a computer keyboard. Therefore, the main task for players is to listen, identify, and react. The novelty of this game was the gamification of a pitch recognition task. In most of the applications reviewed, this task seems to be monotonous; therefore, gamification can engage players' attention. The application was evaluated to find its learning approach by the use of the LM-GM Model. The analysis found that the main game-elements of the game are related to learning elements. For instance, the main task (listen, identify, and react) is a drill-and-practice task, common in the learning approach.

The design of the two prototypes (*Jump'n'Rhythm* and *A Bard's Tale*), furthermore, can be used as a tool to carry on more experiments. These games were designed with simple game-elements in order to not interfere or distract users with complex dynamics. Additionally, they can be adapted with different stimuli and different game-elements.

The Stimuli and Player Behavior in Rhythm Games

The motivation behind writing this thesis was to determine how the different stimuli (emitted in a rhythm video game) affect the interaction between players and music video games. This thesis presented two case studies. However, although results from the case studies were not significant, there are some conclusions and interpretations from the data that can be drawn. In rhythm games, where the task is synchronized with a sequence of events, the auditory or tactile stimuli enhance players' performance more than visual stimuli. In games where a specific finger movement is needed for certain actions, auditory stimuli have been proven to facilitate the association between movement and action. Nevertheless, either with auditory (listening), with visual (observing), or with haptic (acting) stimuli, the body will activate the sensorimotor system to produce an action in an action-perception process.

This thesis found that synchronizing with sequences of events using auditory stimuli is more suitable than visual stimuli. Moreover, accuracy in a synchronization task using tactile stimuli is better than synchronizing with visual stimuli. However, the accuracy using tactile stimuli is slightly less accurate than using auditory stimuli (*H4: Performance of players using tactile stimuli is similar to the performance of players using auditory stimuli*). These results nevertheless need to be interpreted with caution. These conclusions correspond with the action-perception process, where the auditory-motor system dominates temporal perception. Furthermore, the findings of this work suggest the use of tactile stimuli, not only as a decorative element, but also as potential stimuli in tasks that require synchronization to sequences of events.

Consequently, in rhythm video games, designers can change the performance of players by modifying the players' body reactions through the stimuli. Therefore, the use of different game-elements and dynamics that affect the performance of players could be limited.

Additionally, in the first case study (auditory vs. visual stimuli), the musical background of participants (musicians and non-musicians) allowed to assess the relation that music expertise has in participants. According to H2 (*Musicians have a better reaction time than non-musicians using auditory and visual stimuli*), musicians performed better than non-musicians in both conditions (auditory or visual stimuli). Although these data must be interpreted with caution since the results were not statistically significant, some conclusions can be stated.

Therefore, game designers need to know the target group of users, their preferences, skills, and limitations. For instance, designers can adapt the game according to the music experience of players. The design can be modified by changing the game's level of difficulty, not only in the aspect of game-mechanics but also in musical aspects.

In the *Jump'n'Rhythm* prototype, the difficulty level was implemented using hidden game-elements (Prof. Eva Verena suggested). In consequence, two hypotheses emerged. For the case study (auditory vs. visual stimuli) the hypothesis was H1: *Players have a better reaction time using auditory stimuli even in the absence of supportive game-elements*. For the case study auditory vs. tactile, the hypothesis was H5: *Absence of game-elements hinders the performance of the players in both conditions*. The presence of certain game-elements (e.g., platforms, gaps) affects the performance of the players. Nevertheless, these results need to be interpreted with caution because they are not significant. However, it can be concluded that game-elements (platforms) can support or guide the player's actions. The players' attention is focused on these elements (platforms). In games where the aim is to acquire music skills, these guiding elements can interrupt the learning process. Therefore, this thesis proposed that in music games-based learning, guiding elements support players in familiarizing themselves with the game. However, designers must consider increasing the level of challenge by hiding these elements at some point. The aim of hiding the elements is to force the player to pay attention to the stimuli, in this case auditory stimuli.

Emotions and Experience from the Prototypes

Chapter 4 and 5 explore the second question that motivates this project: How do music-based games enhance the players' subjective experience in video games? Music, as well as video games, create a synergy with the person. They originate positive or negative feelings, involve people, and serve as learning or therapeutic tools.

For H3 (*The workload for participants using visual stimuli is higher than participants using auditory stimuli*), this study did not find a significant difference between auditory or visual stimuli regarding the workload. Similarly, for H6 (*The user experience is higher for participants using the auditory stimuli*), this study showed that the results were not statistically significant. Nevertheless, some comments can be stated with caution. Players feel comfortable when a rhythm game has visual elements (e.g., platforms) because they follow the platforms. However, when the elements are hidden, they require more attention to achieve the task. The different designs, where one level is more manageable than the other, can keep the players' interest.

Moreover, at a manageable level, players can learn the dynamic of the game; thus, they can focus on the stimuli for a more complex level. Probably the player experience is more evident in the third study where the pitch prototype was evaluated. Although contrary to expectations, this study did not find a significant difference between the pitch game and the web application. Also, some comments can be stated with caution. From the qualitative data, it can be inferred that the pitch game attracts more players (*H7: Gamification of a pitch recognition task is more attractive to youths than a non-gamified task*). Moreover, according to the literature reviewed, players learn more from a gamified task than from a non-gamified task (*H8: Users learn a task such as pitch recognition through the use of a video game*). The reason is that there are some feelings and emotions that create an optimal state for learning and improve the cognitive skills for learning. For instance, the research found that challenges and skills are vital states in video games-based learning. Thus, depending on the balance between challenge and skills, the emotional state of players changes. The combination of these states vary from apathy to flow or from reaction to anxiety, for instance. However, it is recommended that designers use a model such as the LM-GM Model when they create game-based learning. This type of model allows identifying the learning-elements and its equivalent in game-elements; thus, designers can be sure that their designs will improve learning.

Finally, the surprise factor is essential to keep the attention of players; otherwise, the activity becomes boring, and boredom is related to poor learning. In the case of video games, rewards are used to keep the attention and interest of players (e.g., collect coins, levels, lives, health). Furthermore, this thesis found that after some time of play, players start to exhibit a fatigue effect; thus, a surprise factor can recharge the energy of the player and avoid monotonous feelings.

Future Work

Several questions remain unanswered at present. It might be possible to use a different experimental design to evaluate the subjective experience in future investigations. Moreover, this thesis is the beginning of future research in embodied music cognition in music interactive systems. For the rhythm game *Jump'n'Rhythm*, there are more levels with more difficult rhythm sequences. Therefore, a long-term study is necessary to assess the improvement of rhythm skills. Moreover, a comparison between *Jump'n'Rhythm* and a classic method to acquire rhythm skills is necessary to evaluate the efficacy of using rhythm games to improve rhythm skills. Additionally, the game can be modified to use in case studies to explore the problems related to a deficit of rhythm skills (e.g., dyslexia, dyspraxia). However, there must be specialist (therapist, doctors), and people with rhythm deficit involved in the design.

Regarding the pitch game, (*A Bard's Tale*), a further study with more focus on the auditory-motor theory and muscle memory is suggested. Therefore, different physical interfaces must be explored (e.g., musical keyboard, digital keyboard). Additionally, more levels must be designed to train music intervals, chords, and harmonies.

Moreover, further work is required to establish the relationship between the avatar, the sound, and the body gestures; and how music video games generate empathy through this relationship.

To date, in line with the embodied music cognition, the research was extrapolated to gestures to control and interact with music. For example, the sonification of some yoga positions can engage participants and enhance their experience, rather than only background music. The interaction between body and music in a collaborative activity generates empathy.

Bibliography

Alexandrovsky, D., Alvarez, K., Walther-Franks, B., Wollersheim, J., and Malaka, R. (2016). Jump'n'rhythm: A video game for training timing skills. In Weyers, B. and Dittmar, A., editors, *Mensch und Computer 2016 - Workshopband, Aachen, Germany, September 4-7, 2016*, Aachen, Germany. Gesellschaft für Informatik e.V.

Alvarez-Molina, K., Jöllenbeck, A., and Malaka, R. (2018). Listen and play: Auditory-motor interaction in a bard's tale video game. In Beck, D., Morgado, C. A. L., Pirker, J., Peña-Rios, A., Ogle, T., Richter, J., and Gütl, C., editors, *Immersive Learning Research Network. iLRN 2018.*, volume 840 of *Communications in Computer and Information Science*, pages 198–210. Springer, Cham.

Arnab, S., Lim, T., Carvalho, M. B., Bellotti, F., De Freitas, S., Louchart, S., Suttie, N., Berta, R., and De Gloria, A. (2015). Mapping learning and game mechanics for serious games analysis. *British Journal of Educational Technology*, 46(2):39–411.

Baijal, A., Kim, J., Branje, C., Russo, F., and Fels, D. I. (2012). Composing vibrotactile music: A multi-sensory experience with the emoti-chair. In MacLean, K. and O'Malley, M. K., editors, *2012 IEEE Haptics Symposium (HAPTICS) in Vancouver, BC, Canada*, pages 509–515, Washington, USA. IEEE Computer Society.

Baker, R. S., D'Mello, S. K., Rodrigo, M. M. T., and Graesser, A. C. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive-affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies*, 68(4):223–241.

Bangert, M. and Altenmüller, E. O. (2003). Mapping perception to action in piano practice: A longitudinal dc-eeg study. *BMC neuroscience*, 4, 26.

Bangert, M., Haeusler, U., and Altenmüller, E. (2001). On practice: How the brain connects piano keys and piano sounds. *Annals of the New York Academy of Sciences*, 930:425–428.

Bangert, M., Jürgens, U., Häusler, U., and Altenmüller, E. (2006). Classical conditioned responses to absent tones. *BMC neuroscience*, 7, 60.

Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.*, 59(1):617–645.

Baumann, S., Koeneke, S., Meyer, M., Lutz, K., and Jäncke, L. (2005). A network for sensory-motor integration. *Annals of the New York Academy of Sciences*, 1060:186–188.

Becker, J. (2004). *Deep listeners: Music, emotion, and trancing*. Indiana University Press.

Becker, J. (2022). Music, trancing, and the absence of pain. In Coakley, S., editor, *Pain and Its Transformations: The Interface of Biology and Culture*, pages 166–194. Harvard University Press.

Bégel, V., Di Loreto, I., Seilles, A., and Dalla Bella, S. (2017). Music games: potential application and considerations for rhythmic training. *Frontiers in human neuroscience*, 11, 273.

Bégel, V., Seilles, A., and Dalla Bella, S. (2018). Rhythm workers: A music-based serious game for training rhythm skills. *Music & Science*, 1:1–16.

Bethesda Game Studios (2002). *The Elder Scrolls: Morrowind*. Game [PC and Xbox]. Bethesda Softworks, USA.

Bianchi-Berthouze, N., Kim, W. W., and Patel, D. (2007). Does body movement engage you more in digital game play? and why? In Paiva, A. C., Prada, R., and Picard, R. W., editors, *Affective Computing and Intelligent Interaction. ACII 2007*, Lecture Notes in Computer Science, pages 102–113, Berlin, Heidelberg. Springer Berlin Heidelberg.

Boot, W. R. (2015). Video games as tools to achieve insight into cognitive processes. *Frontiers in Psychology*, 6, 3.

Bootsma, R. and C. W. Van Wieringen, P. (1990). Timing an attacking forehand drive in table tennis. *Journal of Experimental Psychology: Human Perception and Performance*, 16(1):21–29.

Bowey, J. T., Birk, M. V., and Mandryk, R. L. (2015). Manipulating leaderboards to induce player experience. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play*, CHI PLAY '15, page 115–120, New York, USA. Association for Computing Machinery.

Brass, M., Bekkering, H., Wohlschläger, A., and Prinz, W. (2000). Compatibility between observed and executed finger movements: comparing symbolic, spatial, and imitative cues. *Brain and Cognition*, 44(2):124–143.

Brochard, R., Touzalin, P., Després, O., and Dufour, A. (2008). Evidence of beat perception via purely tactile stimulation. *Brain Research*, 1223:59–64.

Brown, E. and Cairns, P. (2004). A grounded investigation of game immersion. In *CHI '04 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '04, page 1297–1300, New York, USA. Association for Computing Machinery.

Buccino, G., Binkofski, F., and Riggio, L. (2004a). The mirror neuron system and action recognition. *Brain and Language*, 89(2):370–376.

Buccino, G., Vogt, S., Ritzl, A., Fink, G. R., Zilles, K., Freund, H.-J., and Rizzolatti, G. (2004b). Neural circuits underlying imitation learning of hand actions: an event-related fmri study. *Neuron*, 42(2):323–334.

Burr, D., Banks, M. S., and Morrone, M. C. (2009). Auditory dominance over vision in the perception of interval duration. *Experimental Brain Research*, 198(1):49–57.

Bytes, P. (2001). *Gothic*. Game [PC]. Egmont Interactive, Germany.

Carey, R. (2015). Game design canvas. [power point presentation]. In *Serious Play Conference, Pittsburgh, PA, USA*.

Castel, A. D., Pratt, J., and Drummond, E. (2005). The effects of action video game experience on the time course of inhibition of return and the efficiency of visual search. *Acta Psychologica*, 119(2):217–230.

Cattaneo, L. and Rizzolatti, G. (2009). The mirror neuron system. *Archives of Neurology*, 66(5):557–560.

Chan, A. and Ng, A. (2012). Finger response times to visual, auditory and tactile modality stimuli. In *2012 International MultiConference of Engineers and Computer Scientists, IMECS 2012*, volume 2196 of *Lecture Notes in Engineering and Computer Science*, pages 1449–1454. Newswood Limited.

Chandler, P. and Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, 8(4):293–332.

Chen, J. L., Penhune, V. B., and Zatorre, R. J. (2008). Moving on Time: Brain Network for Auditory-Motor Synchronization is Modulated by Rhythm Complexity and Musical Training. *Journal of Cognitive Neuroscience*, 20(2):226–239.

Chen, L. and Vroomen, J. (2013). Intersensory binding across space and time: a tutorial review. *Attention, Perception, & Psychophysics*, 75(5):790–811.

Chen, Y., Repp, B. H., and Patel, A. D. (2002). Spectral decomposition of variability in synchronization and continuation tapping: Comparisons between auditory and visual pacing and feedback conditions. *Human Movement Science*, 21(4):515–532.

Cheng, M.-T., She, H.-C., and Annetta, L. (2015). Game immersion experience: its hierarchical structure and impact on game-based science learning. *Journal of Computer Assisted Learning*, 31(3):232–253.

Cheng, M.-T., Su, T., Huang, W.-Y., and Chen, J.-H. (2014). An educational game for learning human immunology: What do students learn and how do they perceive? *British Journal of Educational Technology*, 45(5):820–833.

Chung, S.-M. and Wu, C.-T. (2017). Designing music games and mobile apps for early music learning. In Ma, M. and Oikonomou, A., editors, *Serious Games and Edutainment Applications : Volume II*, pages 57–75. Springer International Publishing, Cham.

Church, D. (2006). Formal abstract design tools. In Salen, K. and Zimmerman, E., editors, *The Game Design Reader: A Rules of Play Anthology*. The MIT Press, Cambridge, Massachusetts.

Clark, A. (1999). An embodied cognitive science? *Trends in Cognitive Sciences*, 3(9):345–351.

Clark, A. and Chalmers, D. (1998). The extended mind. *Analysis*, 58(1):7–19.

Collins, K. (2011). Making gamers cry: Mirror neurons and embodied interaction with game sound. In *Proceedings of the 6th Audio Mostly Conference: A Conference on Interaction with Sound*, AM '11, pages 39–46, New York, NY, USA. ACM.

Collins, K. (2013). *Playing with sound: a theory of interacting with sound and music in video games*. MIT Press.

Cooley, M. (1998). Sound + image in computer-based design: learning from sound in the arts. In Brewster, S., editor, *Proceedings of the 5th International Conference on Auditory Display (ICAD1998)*. British Computer Society.

Cosper, S. M., Lee, G. P., Peters, S. B., and Bishop, E. (2009). Interactive metronome training in children with attention deficit and developmental coordination disorders. *International Journal of Rehabilitation Research*, 32(4):331–336.

Cox, A. (2001). The mimetic hypothesis and embodied musical meaning. *Musicae scientiae*, 5(2):195–212.

CrackedRabbitGaming (October 22, 2006). Donkey konga: Four full combos in 4-player mode, by myself! [video file]. YouTube. `https://youtu.be/OsHK65IOn5M`.

Craighero, L., Bello, A., Fadiga, L., and Rizzolatti, G. (2002). Hand action preparation influences the responses to hand pictures. *Neuropsychologia*, 40(5):492–502.

Csikszentmihalyi, M. (1997). *Finding flow: The psychology of engagement with everyday life*. Basic Books.

Csikszentmihalyi, M., Abuhamdeh, S., and Nakamura, J. (2014). Flow. In *Flow and the Foundations of Positive Psychology: The Collected Works of Mihaly Csikszentmihalyi*, pages 227–238. Springer Netherlands, Dordrecht.

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

Cummins, F., Port, R., et al. (1998). Rhythmic constraints on stress timing in english. *Journal of Phonetics*, 26(2):145–171.

Cytowic, R. E. (1989). *Synesthesia: A Union of the Senses*. Springer New York, New York, NY.

de Castell, S., Jenson, J., and Thumlert, K. (2014). From simulation to imitation: Controllers, corporeality, and mimetic play. *Simulation & Gaming*, 45(3):332–355.

Deci, E. L., Koestner, R., and Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological bulletin*, 125(6):627–700.

Deci, E. L. and Ryan, R. M. (2000). The" what" and" why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological inquiry*, 11(4):227–268.

Denis, G. and Jouvelot, P. (2005). Motivation-driven educational game design: Applying best practices to music education. In *Proceedings of the 2005 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology*, ACE '05, pages 462–465, New York, NY, USA. Association for Computing Machinery.

Derrick, T. R. and Thomas, J. M. (2004). Time series analysis: the cross-correlation function. In Stergiou, N., editor, *Innovative Analyses of Human Movement*, pages 189–205. Human Kinetics Publishers, Champaign, Illinois.

Diederich, A. and Colonius, H. (2004). Bimodal and trimodal multisensory enhancement: Effects of stimulus onset and intensity on reaction time. *Perception & Psychophysics*, 66(8):1388–1404.

Donohue, S. E., Woldorff, M. G., and Mitroff, S. R. (2010). Video game players show more precise multisensory temporal processing abilities. *Attention, Perception, & Psychophysics*, 72(4):1120–1129.

Double Fine Productions (2009). *Brütal Legend*. Game [PS3, Xbox360, PC, iOS, Linux]. Electronic Arts and Double Fine Productions (PC), USA.

Egenfeldt-Nielsen, S. (2011). What makes a good learning game? going beyond edutainment. *ELearn*, 2011(2).

Eisner, E. W. (1991). Rethinking literacy. *Educational Horizons*, 69(3):120–128.

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):202–209.

Escoffier, N., Sheng, D. Y. J., and Schirmer, A. (2010). Unattended musical beats enhance visual processing. *Acta Psychologica*, 135(1):12–16.

Evans, P. (2015). Self-determination theory: An approach to motivation in music education. *Musicae Scientiae*, 19(1):65–83.

Fadiga, L., Fogassi, L., Pavesi, G., and Rizzolatti, G. (1995). Motor facilitation during action observation: a magnetic stimulation study. *Journal of neurophysiology*, 73(6):2608–2611.

Firaxis Games (2005). *Civilization IV*. Game [PC, iOS]. 2K Games (Win) and Aspyr (Mac), USA.

Formanek, R. (1991). Why they collect: Collectors reveal their motivations. *Journal of Social Behavior & Personality*, 6(6):275–286.

Franceschini, S., Trevisan, P., Ronconi, L., Bertoni, S., Colmar, S., Double, K., Facoetti, A., and Gori, S. (2017). Action video games improve reading abilities and visual-to-auditory attentional shifting in english-speaking children with dyslexia. *Scientific reports*, 7(1). 5863.

FreeStyleGames and Exient Entertainment (2009). *DJ Hero*. Game [PS2, PS3, Wii, Xbox 360]. Activision, USA.

Fu, Y. (2016). *Effect of Random Item Dropping on Game Immersion [Unpublished HCI-E MSc Final Project Report]*. UCL Interaction Centre, University College London.

Fullagar, C. J., Knight, P. A., and Sovern, H. S. (2013). Challenge/skill balance, flow, and performance anxiety. *Applied Psychology*, 62(2):236–259.

Gaab, N., Gaser, C., Zaehle, T., Jancke, L., and Schlaug, G. (2003). Functional anatomy of pitch memory—an fmri study with sparse temporal sampling. *NeuroImage*, 19(4):1417–1426.

Gagnon, S., Howe, L., and Waskom, M. (n.d.). Section week 8 - linear mixed models. Retrieved June 20, 2021 from `https://web.archive.org/web/20151112191004/https://web.stanford.edu/class/psych252/section/Mixed_models_tutorial.html`.

Gaijin Games (2010). *Bit Trip Runner*. Game [Wii, PC, iOS, Nintendo3DS, Linux]. Aksys Games (Wii, 3DS) and Gaijin Games (PC), USA.

Gajadhar, B. J., De Kort, Y. A., and Ijsselsteijn, W. A. (2008). Shared fun is doubled fun: Player enjoyment as a function of social setting. In Markopoulos, P., IJsselsteijn, W., and Rowland, D., editors, *Fun and games. Second International Conference, Eindhoven, The Netherlands, October 20-21, 2008. Proceedings*, volume 5294, pages 106–117. Springer, Berlin, Heidelberg, Eindhoven, The Netherlands.

Gallese, V. (2005). Embodied simulation: From neurons to phenomenal experience. *Phenomenology and the Cognitive Sciences*, 4:23–48.

Gallese, V., Fadiga, L., Fogassi, L., and Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, 119(2):593–609.

Gangitano, M., Mottaghy, F. M., and Pascual-Leone, A. (2001). Phase-specific modulation of cortical motor output during movement observation. *Neuroreport*, 12(7):1489–1492.

Garneau, P. A. (October 12, 2001). Fourteen forms of fun. Game Developer. `https://www.gamedeveloper.com/design/fourteen-forms-of-fun`.

Gazzola, V., Rizzolatti, G., Wicker, B., and Keysers, C. (2007). The anthropomorphic brain: the mirror neuron system responds to human and robotic actions. *NeuroImage*, 35(4):1674–1684.

Gebauer, G. and Wulf, C. (1995). *Mimesis: Culture Art Society*. Mimesis: Culture, Art, Society. University of California Press.

Gee, J. P. (2003). What video games have to teach us about learning and literacy. *Comput. Entertain.*, 1(1):20.

Gee, J. P. (2007). *Good video games and good learning*. Peter Lang Mrz.

Glen, S. (2013). Rmse: Root mean square error. from statisticshowto.com: Elementary statistics for the rest of us. `https://www.statisticshowto.com/probability-and-statistics/regression-analysis/rmse-root-mean-square-error/`.

Goh, D. H.-L., Pe-Than, E. P. P., and Lee, C. S. (2015). An investigation of reward systems in human computation games. In Kurosu, M., editor, *Human-Computer Interaction: Interaction Technologies. HCI 2015*, pages 596–607, Cham. Springer International Publishing.

Gonzalez-Grandón, X. and Froese, T. (2018). Grounding 4e cognition in mexico: Introduction to special issue on spotlight on 4e cognition research in mexico. *Adaptive Behavior*, 26(5):189–198.

Gower, L. and McDowall, J. (2012). Interactive music video games and children's musical development. *British Journal of Music Education*, 29(1):91–105.

Greenwald, A. G. (1970). Sensory feedback mechanisms in performance control: With special reference to the ideo-motor mechanism. *Psychological Review*, 77(2):73–99.

Griffin, S. N. (2005). Push. play: An examination of the gameplay button. In *Digital Games Research Conference 2005, Changing Views: Worlds in Play, June 16-20, Vancouver, British Columbia, Canada*, volume 3. Digital Games Research Association.

Hamari, J., Shernoff, D. J., Rowe, E., Coller, B., Asbell-Clarke, J., and Edwards, T. (2016). Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning. *Computers in Human Behavior*, 54:170–179.

Hannon, E. E., Snyder, J. S., Eerola, T., and Krumhansl, C. L. (2004). The role of melodic and temporal cues in perceiving musical meter. *Journal of Experimental Psychology: Human Perception and Performance*, 30(5):956–974.

Hanson, J. V., Whitaker, D., and Heron, J. (2009). Preferential processing of tactile events under conditions of divided attention: Effects of divided attention on reaction time. *NeuroReport*, 20(15):1392–1396.

Hao, W. and Chuen-Tsai, S. (2011). Game reward systems: Gaming experiences and social meanings. In *Proceedings of the 2011 DiGRA International Conference: Think Design Play. 14-17 September, Utrecht, Netherlands*. DiGRA/Utrecht School of the Arts.

Hargreaves, D., Marshall, N., and North, A. (2003). Music education in the twenty-first century: A psychological perspective. *British Journal of Music Education*, 20(2):147–163.

Harmonix (2010). *Dance Central*. Game [Xbox360, XboxOne, Oculus Rift, Oculus Rift S, Oculus Quest]. MTV Games (2010), Microsoft Studios (2011–2014), Oculus Studios (2019), USA.

Harmonix, Budcat Creations, Neversoft., Vicarious Visions, and FreeStyleGames (2005). *Guitar Hero*. Game [PS2, PS3, Wii, PC, Nintendo DS, Android, PS4, Wii U, XboxOne, Xbox360, iOS]. RedOctane (2005–2009) and Activision (2006–), USA.

Harmonix and Pi Studios (2007). *Rock Band*. Game [Xbox360, XboxOne, PS3, PS2, PS4, PSP, Wii, Nintendo DS, iOS]. MTV Games, Mad Catz, Harmonix, Japan.

Hart, S. G. and Staveland, L. E. (1988). Development of nasa-tlx (task load index): Results of empirical and theoretical research. In Hancock, P. A. and Meshkati, N., editors, *Human Mental Workload*, volume 52 of *Advances in Psychology*, pages 139 – 183. North-Holland.

Harter, J.-L. (2002). *Le jeu: essai de déstructuration*. Ats, transversalité, éducatio.

Haueisen, J. and Knösche, T. R. (2001). Involuntary motor activity in pianists evoked by music perception. *Journal of Cognitive Neuroscience*, 13(6):786–792.

Holst, I. (1963). *An ABC of music: a short practical guide to the basic essentials of rudiments, harmony, and form*. Oxford paperback reference. Oxford University Press.

Hove, M. J., Fairhurst, M. T., Kotz, S. A., and Keller, P. E. (2013a). Synchronizing with auditory and visual rhythms: An fmri assessment of modality differences and modality appropriateness. *NeuroImage*, 67:313–321.

Hove, M. J., Iversen, J. R., Zhang, A., and , R. B. H. (2013b). Synchronization with competing visual and auditory rhythms: bouncing ball meets metronome. *Psychological Research*, 77(4):388–398.

Huang, J., Gamble, D., Sarnlertsophon, K., Wang, X., and Hsiao, S. (2013). Integration of auditory and tactile inputs in musical meter perception. In Moore, B. C., Patterson, R. D., Winter, I. M., Carlyon, R. P., and Gockel, H. E., editors, *Basic Aspects of Hearing*, volume 787 of *Advances in Experimental Medicine and Biology*, New York, NY. Springer, New York, NY.

Huss, M., Verney, J. P., Fosker, T., Mead, N., and Goswami, U. (2011). Music, rhythm, rise time perception and developmental dyslexia: perception of musical meter predicts reading and phonology. *Cortex*, 47(6):674–689.

Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C., and Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. *PLOS Biology*, 3(3). e79.

IJsselsteijn, W., Van Den Hoogen, W., Klimmt, C., De Kort, Y., Lindley, C., Mathiak, K., Poels, K., Ravaja, N., Turpeinen, M., and Vorderer, P. (2008). Measuring the experience of digital game enjoyment. In Spink, A., Ballintijn, M., Bogers, N., Grieco, F., Loijens, L., Noldus, L., Smit, G., and Zimmerman, P., editors, *Proceedings of Measuring Behavior 2008, 6th International Conference on Methods and Techniques in Behavioral Research (Maastricht, The Netherlands, August 26-29, 2008)*, pages 88–89. Noldus Information Technology bv.

Indieszero (2005). *Elektropankton*. Game [Nintendo DS and Nintendo DSi]. Nintendo, Japan.

iNiS (2006). *Elite Beat Agents*. Game [Nintendo DS]. Nintendo, Japan.

Iversen, J. R., Patel, A. D., Nicodemus, B., and Emmorey, K. (2015). Synchronization to auditory and visual rhythms in hearing and deaf individuals. *Cognition*, 134:232–244.

Jack, R., McPherson, A., and Stockman, T. (2015). Designing tactile musical devices with and for deaf users: a case study. In Timmers, R., Dibben, N., Eitan, Z., Granot, R., Metcalfe, T., Schiavio, A., and Williamson, V., editors, *Proceedings of ICMEM 2015.*

International Conference on the Multimodal Experience of Music, pages 23–25. Sheffield: The Digital Humanities Institute.

James, W. (1890). *The principles of psychology*, volume 1. Henry Holt and Company.

Jäncke, L. (2006). From cognition to action. In Altenmüller, E., K. J. &. W. M., editor, *Music, Motor Control and the Brain*, pages 25–37. Oxford University Press.

Jennett, C., Cox, A. L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., and Walton, A. (2008). Measuring and defining the experience of immersion in games. *International Journal of Human-Computer Studies*, 66(9):641–661.

Jensenius, A. R. (2007). *Action-sound: Developing methods and tools to study music-related body movement*. PhD thesis, Department of Musicology. University of Oslo, Oslo, Norway.

Jenson, J. and De Castell, S. (2009). From simulation to imitation: New controllers, new forms of play. In *Proceedings of the 2009 DiGRA International Conference: Breaking New Ground: Innovation in Games, Play, Practice and Theory*. Brunel University.

Johnson, D., Klarkowski, M., Vella, K., Phillips, C., McEwan, M., and Watling, C. N. (2018). Greater rewards in videogames lead to more presence, enjoyment and effort. *Computers in Human Behavior*, 87:66–74.

Jones, L. A. and Sarter, N. B. (2008). Tactile displays: Guidance for their design and application. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 50(1):90–111.

Juul, J. (2003). The game, the player, the world: looking for a heart of gameness. In Copier, M. and Raessens, J., editors, *Proceedings of the 2003 DiGRA International Conference: Level Up. 4-6 November, Utrecht, Netherlands*, volume 2, pages 30–45. Universiteit Utrecht.

Juul, J. (2010). *A casual revolution*. MIT Press.

Kato, M. and Konishi, Y. (2006). Auditory dominance in the error correction process: A synchronized tapping study. *Brain Research*, 1084(1):115 – 122.

Kent, S. L. (2001). *The Ultimate History of Video Games: Volume Two: from Pong to Pokemon and beyond... the story behind the craze that touched our lives and changed the world*. Three Rivers Press.

Keysers, C., Kohler, E., Umiltà, M. A., Nanetti, L., Fogassi, L., and Gallese, V. (2003). Audiovisual mirror neurons and action recognition. *Experimental Brain Research*, 153(4):628–636.

King, D. and Delfabbro, P. (2009). Understanding and assisting excessive players of video games: a community psychology perspective. *Australian Community Psychologist*, 21(1):62–74.

King, D., Delfabbro, P., and Griffiths, M. (2010). Video game structural characteristics: A new psychological taxonomy. *International Journal of Mental Health and Addiction*, 8:90–106.

Kirsh, D. and Maglio, P. (1994). On distinguishing epistemic from pragmatic action. *Cognitive science*, 18(4):513–549.

Klimmt, C. (2003). Dimensions and determinants of the enjoyment of playing digital games: A three-level model. In Copier, M. and Raessens, J., editors, *Level up: Digital games research conference*, pages 246–257, Utrech, Netherlands. Faculty of Arts, Utrecht University,.

Kohler, E., Keysers, C., Umilta, M. A., Fogassi, L., Gallese, V., and Rizzolatti, G. (2002). Hearing sounds, understanding actions: action representation in mirror neurons. *Science*, 297(5582):846–848.

Konami (1998). *DanceDance Revolution*. Game [Arcade, Bemani Pocket, Dreamcast, DVD game, GBColor, GameCube, iOS, PC, Android, Nintendo64, PS, PS2, PS3, TV game, Wii, Xbox, Xbox360]. Konami, Nintendo, Disney, Keen, Betson, Japan.

Kort, B., Reilly, R., and Picard, R. W. (2001). An affective model of interplay between emotions and learning: Reengineering educational pedagogy-building a learning companion. In Okamoto, T., Hartley, R., Kinshuk, and Klus, J. P., editors, *Proceedings IEEE International Conference on Advanced Learning Technologies 2001*, pages 43–46. IEEE Computer Society.

Koster, R. (2013). *Theory of fun for game design*. O'Reilly Media, Inc.

Krampe, R. T. (2002). Aging, expertise and fine motor movement. *Neuroscience & Biobehavioral Reviews*, 26(7):769–776.

Kylie, P., Michael, D., Eric, L., and Hay, K. (2011). The nirvana effect: Tapping video games to mediate music learning and interest. *International Journal of Learning and Media*, 3(1):41–59.

Lahav, A., Boulanger, A., Schlaug, G., and Saltzman, E. (2005). The power of listening: Auditory-motor interactions in musical training. *Annals of the New York Academy of Sciences*, 1060(1):189–194.

Lahav, A., Saltzman, E., and Schlaug, G. (2007). Action representation of sound: Audiomotor recognition network while listening to newly acquired actions. *Journal of Neuroscience*, 27(2):308–314.

Lanier, J. (1983). *Moondust*. Game [Commodore 64]. Creative Software, USA.

Latham, A. J., Patston, L. L., and Tippett, L. J. (2013). The virtual brain: 30 years of video-game play and cognitive abilities. *Frontiers in Psychology*, 4, 629.

Law, E. L.-C., Brühlmann, F., and Mekler, E. D. (2018). Systematic review and validation of the game experience questionnaire (geq)-implications for citation and reporting practice. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*, CHI PLAY '18, page 257–270, New York, NY, USA. Association for Computing Machinery.

Leman, M. (2008). *Embodied music cognition and mediation technology*. MIT Press.

Leman, M., Maes, P.-J., Nijs, L., and Van Dyck, E. (2018). What is embodied music cognition? In Bader, R., editor, *Springer Handbook of Systematic Musicology*, pages 747–760. Springer Berlin Heidelberg.

Leman, Marc and Maes, Pieter-Jan (2015). The role of embodiment in the perception of music. *Empirical Musicology Review*, 9(3-4):236–246.

Lewis, J. R., Utesch, B. S., and Maher, D. E. (2015). Measuring perceived usability: The sus, umux-lite, and altusability. *International Journal of Human–Computer Interaction*, 31(8):496–505.

Liebe, M. (2013). Interactivity and music in computer games. In Moormann, P., editor, *Music and Game: Perspectives on a Popular Alliance*, pages 41–62. Springer Fachmedien Wiesbaden, Wiesbaden.

Lipke, M. (September 10, 2017). In-game screen of the game patapon from Patapon Remastered PSN. [online image]. PSNow. `https://www.ps-now.de/ps4/artikel/review/30385/patapon-remastered-psn`.

Liu, L. C. and Hedeker, D. (2006). A mixed-effects regression model for longitudinal multivariate ordinal data. *Biometrics*, 62(1):261–268.

London Studio (2004). *Singstar*. Game [PS2, PS3 and PS4]. Sony Computer Entertainment Europe, Japan.

Maes, P.-J., Buhmann, J., and Leman, M. (2016). 3mo: A model for music-based biofeedback. *Frontiers in Neuroscience*, 10, 548.

Maes, P.-J., Leman, M., Palmer, C., and Wanderley, M. (2014). Action-based effects on music perception. *Frontiers in Psychology*, 4, 1008.

Maes, P.-J., Nijs, L., and Leman, M. (2018). A conceptual framework for music-based interaction systems. In Bader, R., editor, *Springer Handbook of Systematic Musicology*, Springer Handbook, pages 793–804. Springer, Berlin, Heidelberg.

Magezi, D. A. (2015). Linear mixed-effects models for within-participant psychology experiments: an introductory tutorial and free, graphical user interface (lmmgui). *Frontiers in Psychology*, 6, 2.

Malaka, R. (2014). How computer games can improve your health and fitness. In Göbel, S. and Wiemeyer, J., editors, *Games for Training, Education, Health and Sports*, pages 1–7, Cham. Springer International Publishing.

Malone, T. W. (1980). What makes things fun to learn? heuristics for designing instructional computer games. In *Proceedings of the 3rd ACM SIGSMALL Symposium and the First SIGPC Symposium on Small Systems*, SIGSMALL '80, page 162–169, New York, NY, USA. Association for Computing Machinery.

Malone, T. W. and Lepper, M. (1987). Making learning fun: A taxonomy of intrinsic motivations for learning. In Snow, R. E. and Marshall, J., editors, *Aptitude, learning and Instruction*, pages 223–253. Routledge.

Manning, F. and Schutz, M. (2013). Moving to the beat improves timing perception. *Psychonomic Bulletin & Review*, 20(6):1133–1139.

McAuley, E., Duncan, T., and Tammen, V. V. (1989). Psychometric properties of the intrinsic motivation inventory in a competitive sport setting: A confirmatory factor analysis. *Research Quarterly for Exercise and Sport*, 60(1):48–58.

McKernan, B., Martey, R. M., Stromer-Galley, J., Kenski, K., Clegg, B. A., Folkestad, J. E., Rhodes, M. G., Shaw, A., Saulnier, E. T., and Strzalkowski, T. (2015). We don't need no stinkin' badges: The impact of reward features and feeling rewarded in educational games. *Computers in Human Behavior*, 45:299–306.

Menary, R. (2010). Introduction to the special issue on 4e cognition. *Phenomenology and the Cognitive Sciences*, 9(4):459–463.

Microsoft (2010). *Kinect*. Game [Xbox360, XboxOne, PC]. Microsoft, USA.

Missingham, A. (2007). *Why console-games are bigger than rock 'n'roll: What the music sector needs to know and how it can get a piece of the action*. National Foundation for Youth Music, London, UK.

Miyake, Y., Onishi, Y., and Pöppel, E. (2004). Two types of anticipation in synchronization tapping. *Acta neurobiologiae experimentalis*, 64(3):415–426.

Moreno, R. (2010). Cognitive load theory: More food for thought. *Instructional Science*, 38(2):135–141.

Morrison, I. and Ziemke, T. (2005). Empathy with computer game characters : A cognitive neuroscience perspective. In Halle, L., Heylen, D. K., Marsella, S., Pelachaud, C., Wallis, P., and Woods, S., editors, *AISB'05 Convention Social Intelligence and Interaction in Animals, Robots and Agents: 12-15 April 2005, University of Hertfordshire, Hatfield, UK*, volume , s. 73-79. The Society for the Study of Artificial Intelligence and the Simulation of Behaviour.

Musictheory.net (2000). Note ear training [online application]. Musictheory.net. `https://www.musictheory.net/exercises/ear-note`.

Nacke, L. and Lindley, C. A. (2008). Flow and immersion in first-person shooters: Measuring the player's gameplay experience. In *Proceedings of the 2008 Conference on Future Play: Research, Play, Share*, Future Play '08, page 81–88, New York, NY, USA. Association for Computing Machinery.

Namco (2003). *Donkey Konga*. Game [Nintendo GameCube]. Nintendo, Japan.

NanaOn-Sha and Japan Studio (1996). *PaRappa the Rapper*. Game [PS, PSPortable, PS4]. Sony Computer Entertainment, Japan.

Nanayakkara, S., Taylor, E., Wyse, L., and Ong, S. H. (2009). An enhanced musical experience for the deaf: Design and evaluation of a music display and a haptic chair. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, page 337–346, New York, NY, USA. Association for Computing Machinery.

Nguyen, T., Gibbings, A., and Grahn, J. (2018). Rhythm and beat perception. In Bader, R., editor, *Springer Handbook of Systematic Musicology*, pages 507–521. Springer Berlin Heidelberg, Berlin, Heidelberg.

Nijs, L., Coussement, P., Moens, B., Amelinck, D., Lesaffre, M., and Leman, M. (2012a). Interacting with the music paint machine: Relating the constructs of flow experience and presence. *Interacting with Computers*, 24(4):237–250.

Nijs, L., Moens, B., Lesaffre, M., and Leman, M. (2012b). The music paint machine: Stimulating self-monitoring through the generation of creative visual output using a technology-enhanced learning tool. *Journal of New Music Research*, 41(1):79–101.

Nintendo EAD and Nintendo EPD (1985). *Super Mario*. Game [Android, Arcade, GB, GameCube, iOS, Nintendo3DS, Nintendo64, NintendoDS, Nintendo and Super NES, Switch, NEC PC-8801, Sharp X1, Virtual Boy, Wii, Wii U]. Nintendo, Japan.

Nintendo EAD, Intelligent Systems, R. and Namco/Namco Bandai (1992). *Mario Kart*. Game [Wii, Switch, iOS, Android]. Nintendo, Japan and North America.

Northwestern University (2017). Signal processing toolbox: xcorr. Computing Facilities. Electrical and Computer Engineering. Northwestern University. `http://www.ece.northwestern.edu/local-apps/matlabhelp/toolbox/signal/xcorr.html`.

Novak, T. P. and Hoffman, D. L. (1997). Measuring the flow experience among web users. *Interval Research Corporation*, 31(1):1–35.

Novembre, G. and Keller, P. E. (2018). Music and action. In Bader, R., editor, *Springer Handbook of Systematic Musicology*, pages 523–537. Springer Berlin Heidelberg, Berlin, Heidelberg.

O'Regan, J. K. and Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24(5):939–973.

Orfanidis, S. J. (1988). *Optimum signal processing: an introduction*. Collier Macmillan.

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

Overy, K., Nicolson, R. I., Fawcett, A. J., and Clarke, E. F. (2003). Dyslexia and music: measuring musical timing skills. *Dyslexia*, 9(1):18–36.

Parodius8662 (July 4, 2009). Ddrx healing vision -angelic mix- single expert [video file]. Youtube. `https://www.youtube.com/watch?v=dNh75JVsrUM`.

Patel, A. D., Iversen, J. R., Chen, Y., and Repp, B. H. (2005). The influence of metricality and modality on synchronization with a beat. *Experimental Brain Research*, 163(2):226–238.

Perkins, R. E. and Hill, A. B. (1985). Cognitive and affective aspects of boredom. *British Journal of Psychology*, 76(2):221–234.

Peterson, B. M. (2001). Variability of active galactic nuclei. *Advanced Lectures on the Starburst-AGN Connection*, page 3–68.

Pichlmair, M. and Kayali, F. (2007). Levels of sound: On the principles of interactivity in music video games. In Baba, A., editor, *Proceedings of the 2007 DiGRA International Conference: Situated Play*, volume 4. Digital Games Research Association.

Pips (July 10, 2017). How to play parappa the rapper [video file]. Youtube. `https://youtu.be/ZYzs6iHF5Jo`.

Poels, K., de Kort, Y., and Ijsselsteijn, W. (2007). "it is always a lot of fun!": Exploring dimensions of digital game experience using focus group methodology. In *Proceedings of the 2007 Conference on Future Play*, Future Play '07, page 83–89, New York, NY, USA. Association for Computing Machinery.

Polatajko, H. J. and Cantin, N. (2005). Developmental coordination disorder (dyspraxia): An overview of the state of the art. *Seminars in Pediatric Neurology*, 12(4):250–258.

Posner, M. I., Nissen, M. J., and Klein, R. M. (1976). Visual dominance: an information-processing account of its origins and significance. *Psychological Review*, 83(2):157–171.

Prieto, R., Medina-Medina, N., Paderewski, P., and Vela-Gutiérrez, F. L. (2015). Design methodology for educational games based on interactive screenplays. In Camacho, D., Gómez-Martín, M. A., and González-Calero, P. A., editors, *Proceedings 2st Congreso de la Sociedad Española para las Ciencias del Videojuego, Barcelona, Spain, June 24, 2015*, volume 1394 of *CEUR Workshop Proceedings*, pages 90–101. CEUR-WS.org.

Pyramid and SCE Japan Studio (2007). *Patapon*. Game [PSPortable, PS4]. Sony Computer Entertainment, Japan.

Pázhitnov, A. (1984). *Tetris*. Different platforms. Electronika 60, USSR.

Quantic Dream (2010). *Heavy Rain*. Game [PS3, PS4, PC]. Sony Computer Entertainment and Quantic Dream (PC), Japan and France.

Ramos Mejia, C. M. (1947). La dinámica del violinista. *Ricordi Americana SAEC, Buenos Aires*.

React (August 2, 2015). React gaming - Batman: Arkham knight [video file]. Youtube. `https://youtu.be/y1EmPZFG5wA`.

Recanzone, G. H. (2003). Auditory influences on visual temporal rate perception. *Journal of Neurophysiology*, 89(2):1078–1093.

RedFox (March 30, 2016). Gothic 3 - stupid death sounds [video file]. YouTube. `https://www.youtube.com/watch?v=Yhbg6_Ip010`.

Reeves, B. (September 13, 2011). In-game screen of the game Bit Trip Runner from Bit.Trip Complete Review. [online image]. Gameinformed. `https://www.gameinformer.com/games/bittrip_complete/b/wii/archive/2011/09/13/gaijin-games-reinvents-the-classic-gaming-wheel-with-its-bit-trip-series.aspx`.

Repp, B. H. (2005). Sensorimotor synchronization: a review of the tapping literature. *Psychonomic Bulletin & Review*, 12(6):969–992.

Repp, B. H. and Penel, A. (2002). Auditory dominance in temporal processing: new evidence from synchronization with simultaneous visual and auditory sequences. *Journal of Experimental Psychology: Human Perception and Performance*, 28(5):1085–1099.

Repp, B. H. and Penel, A. (2004). Rhythmic movement is attracted more strongly to auditory than to visual rhythms. *Psychological Research*, 68(4):252–270.

Repp, B. H. and Su, Y.-H. (2013). Sensorimotor synchronization: a review of recent research (2006–2012). *Psychonomic Bulletin & Review*, 20(3):403–452.

Rizzolatti, G. (2005). The mirror neuron system and its function in humans. *Anatomy and embryology*, 210(5):419–421.

Roche, R., Viswanathan, P., Clark, J. E., and Whitall, J. (2016). Children with developmental coordination disorder (dcd) can adapt to perceptible and subliminal rhythm changes but are more variable. *Human Movement Science*, 50:19–29.

Rocksteady Studios, WB Games Montréal, Splash Damage, Human Head Studios, NetherRealm Studios, Armature Studio, and Turbine, Inc. (2009). *Batman: Arkham*. Game [Android, iOS, PC, Nintendo3DS, PS3, PS4, PSVita, Wii U, Xbox360, Xbox-One]. U.K.: Eidos Interactive (2009), Japan: Square Enix (2009), U.S.: Warner Bros. Interactive Entertainment (2009–present).

Rosemann, S., Altenmüller, E., and Fahle, M. (2016). The art of sight-reading: Influence of practice, playing tempo, complexity and cognitive skills on the eye–hand span in pianists. *Psychology of Music*, 44(4):658–673.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological review*, 110(1):145–172.

Schiavio, A. and Altenmüller, E. (2015). Exploring music-based rehabilitation for parkinsonism through embodied cognitive science. *Frontiers in Neurology*, 6, 217.

Schiavio, A. and van der Schyff, D. (2018). 4e music pedagogy and the principles of self-organization. *Behavioral Sciences*, 8(8). 72.

Schiefele, U. (2001). The role of interest in motivation and learning. In Collins, J. M. and Messick, S., editors, *Intelligence and personality: Bridging the gap in theory and measurement*, pages 163–193. Lawrence Erlbaum Associates, Mahwah, NJ.

Schmidt, R., Richardson, M. J., Arsenault, C., and Galantucci, B. (2007). Visual tracking and entrainment to an environmental rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, 33(4):860–870.

Schrader, C. and Bastiaens, T. J. (2012). The influence of virtual presence: Effects on experienced cognitive load and learning outcomes in educational computer games. *Computers in Human Behavior*, 28(2):648–658.

Seel, J. (1997). Plugged in, spaced out, and turned on: Electronic entertainment and moral mindfields. *The Journal of Education*, 179(3):17–32.

Seifert, U. (2008). The co-evolution of humans and machines: A paradox of interactivity. In Seifert, U., Kim, J. H., and Moore, A., editors, *Paradoxes of Interactivity. Perspectives for Media Theory, Human-Computer Interaction, and Artistic Investigations*, pages 8–25. Transcript Verlag, Bielefeld, Germany.

Seim, C., Reynolds-Haertle, S., Srinivas, S., and Starner, T. (2016). Tactile taps teach rhythmic text entry: Passive haptic learning of morse code. In *Proceedings of the 2016 ACM International Symposium on Wearable Computers*, ISWC '16, page 164–171, New York, NY, USA. Association for Computing Machinery.

Shams, L., Kamitani, Y., and Shimojo, S. (2000). What you see is what you hear. *Nature*, 408(6814):788.

Shams, L., Kamitani, Y., and Shimojo, S. (2002). Visual illusion induced by sound. *Cognitive Brain Research*, 14(1):147–152.

Shrikanth, S. (n.d). Chapter 17: Mixed effects modeling. A Language, not a Letter: Learining Statistics in R. `https://ademos.people.uic.edu/Chapter17.html`.

Sneha, H. (January 4, 2017). Understanding correlation. All About Circuits. `https://www.allaboutcircuits.com/technical-articles/understanding-correlation/`.

SPD, N. (2006). *Rhythm Tengoku*. Game [Game Boy Advance Arcade]. Nintendo, Japan.

Squire, K. (2008). Video-game literacy: A literacy of expertise. In Coiro, J. and Knobel, Michele, L. D. J. L., editors, *Handbook of Research on New Literacies*, pages 635–670. Routledge, New York, USA.

Stenberg, A. and Cross, I. (2019). White spaces, music notation and the facilitation of sight-reading. *Scientific Reports*, 9, 5299.

Strafella, A. P. and Paus, T. (2000). Modulation of cortical excitability during action observation: a transcranial magnetic stimulation study. *NeuroReport*, 11(10):2289–2292.

Strati, A. D., Shernoff, D. J., and Kackar, H. Z. (2011). Flow. In Levesque, R., editor, *Encyclopedia of Adolescence*, pages 1050–1059. Springer, New York, NY.

Talbot, W. H., Darian-Smith, I., Kornhuber, H. H., and Mountcastle, V. B. (1968). The sense of flutter-vibration: comparison of the human capacity with response patterns of mechanoreceptive afferents from the monkey hand. *Journal of Neurophysiology*, 31(2):301–334.

Tan, S. (September 5, 2012). Understanding cognition: All of the skills necessary for us to successfully function in everyday life. Psychology Today. `https://www.psychologytoday.com/intl/blog/wise/201209/understanding-cognition`.

The Editors of Encyclopaedia Britanica (2018, February 15). Staff. `https://www.britannica.com/art/staff`.

Tierney, A. and Kraus, N. (2015). Evidence for multiple rhythmic skills. *PLOS ONE*, 10(9). e0136645.

Tierney, A., White-Schwoch, T., MacLean, J., and Kraus, N. (2017). Individual Differences in Rhythm Skills: Links with Neural Consistency and Linguistic Ability. *Journal of Cognitive Neuroscience*, 29(5):855–868.

Tobias, E. (2012). Let's play! learning music through video games and virtual worlds. In McPherson, G. E. and Welch, G. F., editors, *Oxford Handbook of Music Education*, volume 2, pages 531–548. Oxford University Press.

Trainor, L. J., Gao, X., jiang Lei, J., Lehtovaara, K., and Harris, L. R. (2009). The primal role of the vestibular system in determining musical rhythm. *Cortex*, 45(1):35–43. Special Issue on "The Rhythmic Brain".

Tremoulet, P. D. and Feldman, J. (2000). Perception of animacy from the motion of a single object. *Perception*, 29(8):943–951.

Tresilian, J. R. (1994). Perceptual and motor processes in interceptive timing. *Human Movement Science*, 13(3):335 – 373.

Truitt, F. E., Clifton, C., Pollatsek, A., and Rayner, K. (1997). The perceptual span and the eye-hand span in sight reading music. *Visual Cognition*, 4(2):143–161.

Varela, F. J., Thompson, E., and Rosch, E. (2017). *The embodied mind: Cognitive science and human experience*. MIT Press.

Video Games Museum (2006). In-game screen of the game elite beat agents on nintendo ds [online image]. Video Games Museum. `https://www.video-games-museum.com/en/game/Elite-Beat-Agents/81/5/3153`.

VirtualCode.es (June 17, 2018). Music tutor sight read [mobile application]. Google Play Store. `https://play.google.com/store/apps/details?id=es.virtualcode.musictutor.sightread_lite&hl=es_MX`.

Wada, Y., Kitagawa, N., and Noguchi, K. (2003). Audio–visual integration in temporal perception. *International Journal of Psychophysiology*, 50(1):117–124.

Ward, J., Tsakanikos, E., and Bray, A. (2006). Synaesthesia for reading and playing musical notes. *Neurocase*, 12(1):27–34.

Weizsäcker, V. (1940). *Der Gestaltkreis: Theorie der Einheit von Wahrnehmen und Bewegen*. Georg Thieme.

Welch, G. F., Howard, D. M., and Rush, C. (1989). Real-time visual feedback in the development of vocal pitch accuracy in singing. *Psychology of Music*, 17(2):146–157.

Welch, R. B. (1999). Meaning, attention, and the "unity assumption" in the intersensory bias of spatial and temporal perceptions. In Aschersleben, G., Bachmann, T., and Müsseler, J., editors, *Cognitive Contributions to the Perception of Spatial and Temporal Events*, volume 129 of *Advances in Psychology*, pages 371–387. North-Holland.

Welch, R. B. and Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*, 88(3):638–667.

West, G. L., Stevens, S. A., Pun, C., and Pratt, J. (2008). Visuospatial experience modulates attentional capture: Evidence from action video game players. *Journal of Vision*, 8(16):13–13.

WholeTone-Games (July 10, 2018). Pitch perfect ear training [pc application]. STEAM Store. `https://store.steampowered.com/app/888710/Pitch_Perfect_Ear_Training/`.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, 9(4):625–636.

Woodruff Carr, K., Fitzroy, A. B., Tierney, A., White-Schwoch, T., and Kraus, N. (2017). Incorporation of feedback during beat synchronization is an index of neural maturation and reading skills. *Brain and Language*, 164:43–52.

Wristen, B. (2005). Cognition and motor execution in piano sight-reading: A review of literature. *Update: Applications of Research in Music Education*, 24(1):44–56.

Yuan, B. and Folmer, E. (2008). Blind hero: Enabling guitar hero for the visually impaired. In *Proceedings of the 10th International ACM SIGACCESS Conference on Computers and Accessibility*, Assets '08, page 169–176, New York, NY, USA. Association for Computing Machinery.

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.

List of Figures

List of Tables

Appendices

AppendixA

Previews Definitions of Game

As a Keynote Jesper Juul presented at the Level Up conference in Utrecht his talk ”The game, the player, the world: looking for a heart of gameness” (available in Conference Proceedings (Juul, 2003)). Juul concluded a definition of a game based on previews definitions and which cover six main points:

1. Rules: Games are rule-based.
2. Variable, quantifiable outcome: Games have variable, quantifiable outcomes.
3. Value assigned to possible outcomes: That the different potential outcomes of the game are assigned different values, some being positive, some being negative.
4. Player effort: That the player invests effort in order to influence the outcome. (I.e. games are challenging.)
5. Player attached to the outcome: That the players are attached to the outcomes of the game in the sense that a player will be the winner and “happy” if a positive outcome happens, and loser and “unhappy” if a negative outcome happens.
6. Negotiable consequences: The same game [set of rules] can be played with or without real-life consequences.

TABLE A.1: *Table taken from Juul (2003). Table shows previews definitions of game by different authors.*

Source	Definition
Johan Huizinga 1950, p.13.	[...] a free activity standing quite consciously outside ″ordinary″ life as being ″not serious″, but at the same time absorbing the player intensely and utterly. It is an activity connected with no material interest, and no profit can be gained by it. It proceeds within its own proper boundaries of time and space according to fixed rules and in an orderly manner. It promotes the formation of social groupings which tend to surround themselves with secrecy and to stress their difference from the common world by disguise or other means.
Roger Caillois 1961, p.10-11.	[...] an activity which is essentially: Free (voluntary), separate [in time and space], uncertain, unproductive, governed by rules, makebelieve.
Bernard Suits 1978, p. 34.	To play a game is to engage in activity directed towards bringing about a specific state of affairs, using only means permitted by rules, where the rules prohibit more efficient in favor of less efficient means, and where such rules are accepted just because they make possible such activity.
Avedon & Sutton Smith 1981, p.7.	At its most elementary level then we can define game as an exercise of voluntary control systems in which there is an opposition between forces, confined by a procedure and rules in order to produce a disequilibria outcome.

Source	Definition
Chris Crawford 1981, chapter 2.	I perceive four common factors: representation ["a closed formal system that subjectively represents a subset of reality"], interaction, conflict, and safety ["the results of a game are always less harsh than the situations the game models"].
David Kelley 1988, p.50.	a game is a form of recreation constituted by a set of rules that specify an object to be attained and the permissible means of attaining it.
Katie Salen & Eric Zimmerman 2003, p.96.	A game is a system in which players engage in an artificial conflict, defined by rules, that results in a quantifiable outcome.

AppendixB

Color - Key Association

To each pitch correspond a key and a color. The keys were chosen following the same disposition as piano keys (in ascending order from left to right). The aim was to simulate the piano keys. The colors were selected according to the gamut of color that the synaesthesia [1] has. However, from this gamut, there are several color palettes; thus, the prototype uses the most representative.

TABLE B.1: *The table shows the corresponding element between pitch, key, and color implemented in the pitch prototype.*

Pitch	**Key**	**Color**
C	1	Blue
D	2	Green
E	3	Yellow (light)
F	4	Yellow (dark)
G	5	Red
A	6	Magenta
B	7	Purple

[1] Synaesthesia is a condition in which perceptual experiences are induced by stimuli that are not normally associated with such an experience (Ward et al., 2006). For instance, persons see colors while listening to music.

AppendixC

Table for Pitch, Key, and Color

Fragment of the table of King et al. (2010) regarding the five-feature model of video game structural characteristic. The table shows only the feature of rewards and punishment.

TABLE C.1: *The table displays the different features of rewards and punishment with some examples.*

Feature Type	**Sub-features**	**Example**
Reward	General reward type features	Experience points, bonuses
and	Punishment features	Losing life, restarting a level
punishment	Meta-game reward features	Xbox 360 Achievement points
	Intermittent reward features	Increasing difficulty of levels
	Negative reward features	Gaining health, repairing items
	Near miss features	Difficult ”boss” at end of level
	Event frequency features	Unlimited replayability of game
	Event duration features	MMORPGs have no endpoint
	Payout intervals features	Rewarded instantly for playing

AppendixD

Music video game interaction patterns

Table shows some examples of video games and its corresponding interaction patterns according with the principles of interactivity in music video games propose by Pichlmair and Kayali (2007).

TABLE D.1: *Music video game interaction patterns (Pichlmair and Kayali, 2007)*

Game	**1**	**2**	**3**	**4**	**5**	**6**	**7***
Rez	x	x	-	-	x	x	x
Otocky	?	x	-	x	x	-	-
Sim Tunes	x	x	-	x	x	-	-
Electroplankton	x	x	-	x	x	x	x
Vin Ribbon	-	-	x	-	-	x	x
Rhythm Tengoku	-	-	x	-	-	-	-
Elite Beat Agents	-	-	x	-	-	-	-
Guitar Hero	-	-	x	-	-	-	x
FreQuency	-	-	x	-	x	-	-

* quantisation (1), sound agents (2), rhythm action (3), active score (4), free-form play (5), synaesthesia (6), play as performance (7)

Acknowledgments

I acknowledge DAAD (Deutscher Akademischer Austauschdienst) for funding my doctoral project.

I am extremely grateful to my Ph.D. supervisor Prof. Rainer Malaka who supported me during all my doctoral project and made this work possible. He kindly agreed to supervise my project and was always available to help me whenever I needed it.

I would like to thank Prof. Christian Gütl for always being willing to help me and appraising this thesis. Additionally, I would like to express gratitude to Prof. Eva Verena Schmid for her treasured support and advice. I am incredibly grateful to Nancy Barrington for proofreading this thesis, your feedback and advice, and all the time that you dedicated to me.

I am incredibly grateful to Joshua Wollersheim, Aaron Jöllenbeck, and Tim Kratky for their collaboration and contribution to this project. I would like to thank Rosa Haroon and Production Partner Magazine for their help recruiting participants.

I am deeply thankful to my colleagues at the Digital Media Lab: Svenja, Tanja, Thomas M., Dirk, Georg, Johannes, Thomas F., Himanshu, Insa, Jan, Supara, Mark, Mehrdad, Denis, Michael, and all the members for creating a friendly working atmosphere. I would like to thank Sebastian, Ambreen, and Susanne for sharing the office with me. Also, I am extremely grateful to Dimi, Robert, Anke, Nina, and Irmgard for your support, advice, and feedback.

I am deeply thankful to Ameris, Andrea, Irene, Liss, Luis Carlos, Sara, Alex, Marco, Sophie, Alejandra, Alfredo, Valentina, Buse, Carme, Stefan, Samy and to Daniel Diethei, Amir, Mark, Rosa, and all friends that made my time in Germany one of the happiest experience of my life.

Thank God for keeping my family, friends, DM Lab members, and me healthy.

I am greatly thankful to my mom for always being there, supporting me with all my crazy things, and trusting me. Thanks to my sister Lucero because she is always my academic example, you are the best biologist. And thanks to my nephew Tadeo for existing, play and laugh with me. Thank God because I still have my grandpa Pato.

I would like to thank my friend Alinne for cheering me up when I was desperate with the thesis. Thanks to my friends Isela, Yaya, Paty, Miriam, Angel, and Rosy to always support me.

Last and foremost, I am extremely grateful to my husband Diego for cheering me up for his support, patience, advice, and most importantly, his love. Thank you, Diego, for being there, sharing this experience with me, and academically growing up together.

> *"This thesis is dedicated to my Matilde; you were the most important being for me. Thanks for always being there, waiting for me, and giving happiness and emotion to my life. This is your last thesis; you did it!"*

Erklärung zur Dissertation

Ich versichere, dass ich die von mir vorgelegte Dissertation selbstständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit - einschließlich Tabellen, Karten und Abbildungen -, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie - abgesehen von unten angegebenen Teilpublikationen - noch nicht veröffentlicht worden ist sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde.

Die Bestimmungen dieser Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Dr. Rainer Malaka betreut worden.

Bremen, 2021.

Teilpublikationen:

Katya Alvarez-Molina, Anke V. Reinschluessel, Tim Kratky, Martin Scharpenberg, and Rainer Malaka. Can You Feel the Rhythm? Comparing Tactile and Auditory Stimuli in a Rhythm Video Game. Article Submitted to ACM Transactions on Applied Perception (TAP). Submitted.

Alvarez-Molina, K.A., Jöllenbeck A., Malaka R. (In press) Listen and Play: Auditory-Motor Interaction in a Bard's Tale Video Game. In: Beck D. et al. (eds) Immersive Learning Research Network. iLRN 2018. Communications in Computer and Information Science. Springer, Cham.

Alexandrovsky, D., Alvarez, K., Walther-Franks, B., Wollersheim, J. & Malaka, R., (2016). Jump'n'Rhythm: A Video Game for Training Timing Skills. In: Weyers, b.& Dittmar, A. (Hrsg.), Mensch und Computer 2016–Workshopband. Aachen: Gesellschaft für Informatik e.V.