Professional Thesis
Report

Impact of the synchronisation of ambient visual effects to the musical soundtrack on the player's experience in non-realistic solo games

Pérot Alexandre

Tutor: Stephane Cugnet

GOBELINS L'ÉCOLE DE L'IMAGE le cnam

Abstract

The synchronization of sound and visual is an old concept in audiovisual history that is noticeably relevant in today's entertainment industry. Its application to video games outside of rhythm game is possible by tampering with the animations of elements that do not affect the gameplay like background props. This subject of this work was to determine whether or not synchronization of these ambient visual effects had an influence on player experience in single-player non-realistic games. An classical experiment protocol was conducted which consisted in a playthrough of a game with (test) and without (control) synchronization to the music, followed by a questionnaire. For this purpose, a side-scrolling 2D game was developed and strategies to efficiently implement this synchronization were implemented. The results of the experiment did not show any measurable difference between the control and test groups. By taking into account the limitations and biases of the experiment protocol, these findings were not deemed enough to strictly and universally infirm the hypothesis and ideas for a new experimental protocol were developed.

Résumé

La synchronisation du son et du visuel et un concept ancien dans l'historique de l'audiovisuel qui est cependant particulièrement pertinent dans l'industrie du divertissement actuelle. Son application au jeu vidéo en dehors de la catégorie des jeux de rythme est possible en jouant sur les animations des éléments qui n'affectent pas le gameplay comme les objets du décor. Le sujet de ce travail était de déterminer si la synchronisation des effets visuels ambient ont une influence sur l'expérience de jeu des joueurs dans des jeu solo graphiquement non réalistes. Le protocole expérimental classique a été mis en place avec une séance de jeu avec (test) ou sans (contrôle) synchronisation à la musique, suivie d'un questionnaire. Pour ce faire, un jeu side-scroller 2D a été développé et des méthodes pour implémenter de façon efficace cette synchronisation ont été implémentées. Le résultat de l'expérience n'a pas montré de différence mesurable entre les groupes de contrôle et de test. Après considération des biais et limites du protocole expérimentale mis en œuvre, les résultats n'ont pas été considérés comme suffisant pour infirmer l'hypothèse et des pistes pour un nouveau protocole expérimental ont été creusées.

Acknowledgment

I would like to thank my tutor Stephane Cugnet for his help and guidance during this exercise.

I am also grateful towards the pedagogical team of the formation for their supervision and responsiveness.

Finally, I am thankful to all the people that took the time to participate in this experiment or that gave me feedback about its design.

Contents

1 Introduction

The superposition of images and sounds is an incredibly hot topic in the current entertainment landscape. Most of the big social media applications are developing their own version of Tiktok/Douyin audiovisual combo to try to reach their impressive engagement. The popularity of the music and video association is being leveraged by entertainment companies: musical section in movies, trailers, viralization of choreographies pop music...

In game communication, synchronization of visuals to the music is featured heavily, with meticulously crafted trailers where action and rhythm flow in pair (Genshin Impact [1] character's teaser for example). But what is interesting to inquire about is whether or not this should be a subject of interest for the interactive experience itself. The first examples that come to mind are rhythm games, although they are quite a specific category of games whose variety is inherently limited by the difficulty of the rhythm detection and input tasks themselves [2]. However, beyond this first answer, the question of introducing synchronization of visual elements to the music in non rhythm games is compelling. In fact, games are usually composed of a multitudes of interacting entities which all have motion of their own and therefore represent many opportunities to implement this paradigm. In particular, a significant amount of these animated elements are not immediately linked to the main action (i.e the player's action or gameplay). For example, there could be movement of vegetation, light effects like water caustics or even particle systems like dust or bubbles. These elements could be controlled by the music without altering the gameplay loop.

Therefore, we will try to answer whether or not the synchronization of the aforementioned ambient visual effects to the musical soundtrack of a video game has an impact on the user experience of the player, especially their immersion, a notion that is going to be pinpointed later in this paper. Because synchronization of these to the music is not a real phenomenon, we will focus in this work on games with non-realistic visual art direction. Also, we will limit ourselves in this work to single player experiences as the addition of the social dimension can add complexity when it come to synchronization and it would be complex considering the scope of this exercise to include this dimension in an experiment.

2 Literature review

2.1 Sound and visual synchronization in audio-visual art

2.1.1 Definition

The first challenge is trying to define what the notion of synchronization can mean for visual and sound, especially in the context of video games. Synchronization of two data streams (in this case audio and video) is a concept that is stronger that a simple non independence of the two. For example, if we imagine a game where footstep sounds are triggered within a random delay of maximum 2 seconds whenever the player's feet visually hit the ground, we can not say that the audio is independent from the visual (when the player stops walking, the sound will stop too after a few seconds). However, the sound and the visuals would generally be considered to be poorly synchronized, if at all.

According to the Oxford dictionary, "synchronization" means "the fact of happening at the same time or moving at the same speed as something else". Although quite appropriate for this work's subject, some aspects of this definition need to be refined. Synchronization is described as a black or white phenomenon (either two events are synchronized or not). This is not in line with a lot of paper's conception of the notion (for example [3]) which tend to consider synchronization as "gradual" descriptor. Moreover, whereas this all-or-nothing vision can make sense in physics studies, this simple definition does not address in our case the subjectivity of human perception and the variability of the perception of synchronicity when dealing with multiple channels of information. Indeed, [4] shows that the Point of Subjective Synchrony (when someone will consider two signals to be synchronous) can vary from one person to another and this threshold may be even higher when considering events with "looser" notion of timing (like animated motion).

2.1.2 History of synchronization in film and animation

A first interesting starting point for this subject is to understand the relationship between the sound and the visual counterparts in film and animation.

Contrary to what the name may indicate, silent film relates more to the technological constraint of having to use a live band to produce a soundtrack than to the actual absence of sound during the projection of the movie. These tracks were often adapted to the events of the movie using cue sheet which served as a basis for significant improvisation [5].

Don Juan [6], produced by Warner Bros, is considered to be the first film with an embedded soundtrack. When it comes to animation, Ko-Ko Song Car-Tunes [7], a series of songs with a bouncing ball to keep track of the lyrics, were the first animations with sound. However, it is Walt Disney's Steamboat Willie [8] in 1928 that launched the genre popularity. According to [9], the success of this piece is due to how sound was given the foreground in terms of importance. In the piece, the soundtrack is heavily correlated and synchronized to the animation. Disney and Wilfred Jackson worked in tandem to make musically accurate control sheet to indicate action constraints to the score. Indeed, as Disney used his knowledge about the dynamics required to animate the action, Jackson converted it so it could sync with the music's tempo and subdivisions.

In the following Disney animation films, director and composer worked together to plan out the temporality of the movies and create so called "Dope sheets" [9]. This approach was a bold positioning compared to the rest of the field. According to [10], there was a clear distinction between action and animated films as the latter gave much less attention to the music track due to a less serious image. Indeed, as animation was considered amusement material, proper planning of the music score with 'cue sheets' for example was not common practice.

An explanation of the profitability and pertinence of this investment in sound and visual synchronization in animation can be found in the role attributed to music by Michel Chion as an aid to "the apprehension of visual movements" [9]. On top of this, Carlos Baena, senior animator who worked on multiple Pixar movies, highlights timing as a common denominator of music and animation and promotes a synesthesic animation paradigm of considering objects as additional music instruments [11].

From the impact and influence of Disney's work emerged the general term "mickeymousing" (synonym for synchronized scoring) to refer to the synchronizing in film or animation of the screen action with the music track. The Informer [12] and Alexandre Nevski [13] are prime example of extreme mickeymousing in film. Uurnog [14] is an example of intense mickeymousing in video games with its algorithmic audio system that closely follows the gameplay and action.

It is interesting to look at the complaints against mickeymousing which has acquired a negative connotation associated with extreme examples of synchronization. Chuck Jones for example criticizes an excessive attention and time devoted by cartoon musicians to achieving perfect synchronizing instead of focusing on the originality and pertinence of their work. Some complaints are more extreme with it being considered vulgar or tacky [15].

Synchronization is also a tool that can be diverted, especially in the horror genre. For example, Spadoni [16] attributes part of the success of the Dracula [17] and Frankenstein [18] movies to their exploitation of the technical limitation of sound-visual synchronization at the time to increase the creepiness of the scenes. This relates to verified phenomena as [3] proved that lack of synchronization between a virtual character's voice and their lips movement was correlated with an increased perceived scariness of the character.

2.1.3 Synchronization in games

In games, the notion of "action" is closely linked to the gameplay and player interaction. Action and visuals are tightly linked in the overwhelming majority of games, as the game logic and graphics are very intertwined through the sharing of spatial transform data in the game engines (meshes and collider corresponding to the same object depends on the same location, rotation and scale data most of the time).

In most cases, sound follows the visual action. For example, hit sounds are triggered by gameplay events or footsteps sounds by animation events. However, there are also some dynamics that work in the other direction. Lip-sync animations for example are planned the other way around, with the text being written first and the face then being animated consequently (sometimes with the help of procedural tools, like in Mass Effect [19] [20]).

In video games, dynamic music refers to a music track that adapts and reacts to in-game situations and actions. While in some games (e.g Uurnog [14] or Sword of Symphony [21]), dynamic music is tightly linked to the action, in most cases, dynamic music has a rather big timescale (blend and fades that take multiple seconds) to the point where it does not exactly fit the notion of synchronization.

A specific game genre does the inverted process and put music in the foreground. In rhythm games gameplay (thus action) becomes subjected to the music's rhythm to some extent. Synchronization takes the form of a gameplay element in most of these games (e.g osu! [22]). Interestingly, the author of Crypt of the NecroDancer [23] explained how the synchronization aspect of the gameplay had to be faked in this case as the challenge of synchronization was too hard to add any additional tasks on top [2].

Although not synchronization per se, the combination of a game's visual content with specific music was proven to have significant effects on immersion [24] and emotion perception [25].

2.1.4 Synchronization workflows

An interesting parallel between cinema and games can be identified where the main action becomes player actions. In audiovisual art, there are two ways to enforce synchronization, which have an impact on the final product and are chosen depending on pipeline and artistic considerations:

- **adapting the visual, to match musical events and dynamics.** In this category, we have rhythm games, musicals, opera, dance etc... In this workflow, there is usually a back and forth between sound and visual/animation/gameplay teams rather than a strict dominance of the sound.

- **adapting the music to match visual action.** This is the usual workflow used when speaking about mickeymousing [9] but for games, this most commonly takes the form of dynamic music.

The study of extreme examples of mickey mousing are actually not what is proposed by this work. Indeed, in theses instances, the entanglement of music and visual put a lot of constraints on both ends (because back and forth is inevitable since the music composers can't predict a rhythm that would perfectly match gameplay intent) which greatly increases the production costs and limits the creative freedom of one (if not both) aspects. Moreover, it is interesting to note that in audiovisual art, frequent synchrony seems to be of lower quality when sound is subjugated to the music: because every movement is associated with a musical event, for the musical track to not be completely alienated by the other and keep some of its integrity, there is some leniency in the synchronization and delays start to appear (this can be heard in The Informer [12] for example).

In addition, in games that do not put the music track in the foreground, gameplay can't be alienated to the music either without altering the game's identity. Generally, if a player's action is meant to be triggered in a very small delay, synchrony with most music track's divisions/subdivisions is not possible or will not be noticed.

However, there is a specificity in video games that is not as present in non interactive audiovisual media: a clearer distinction between the gameplay and the world. Indeed, the game's environment usually exists on its own with a clearer separation from the gameplay. This could be due to various reasons:

- a player can stay longer in a given environment (e.g a puzzle game or an arcade game) so it needs to have more depth and engaging elements than a movie's background

- the shot must be interesting when there is no action (considering the action depends on the player input), especially with a dynamic camera that makes it harder to direct the shots.

This can explain why the game environment is where synchronization to the music can be seen in non rhythm games. Mario Kart 8 [26] and Mario 3D World [27] are nice examples. In these games, gameplay is not synchronized to the music. However, in some levels, the environment (which has no impact on the gameplay) is synchronized to the music. For example, the toads in the Mario Kart Stadium [26] track or the flowers in the first level of [27] both hop in synchrony to the game music's beat.

It is interesting to note that in very nervous games, rhythms appear in the gameplay. Whether or not these can be linked with the music would be an extensive study subject out of the scope of this work.

2.2 Player immersion

Identifying in 2004 a lack of precise definition for the notion of immersion in video games and virtual reality related work, [28] associated the term with the degree of involvement of the user, which they subdivided in three increasing levels: engagement (the player is interested and wants to continue the experience), engrossment (the player is emotionally invested is the game and has suspended they disbelief) and total immersion (the player's feelings and thoughts are being influenced exclusively by the game).

The latter level has been equated with the notion of presence [29], which has been defined by [30] as the "illusion of nonmediation". The gradual definition of immersion was supported by [31] but its intermingling with presence was questioned. More generally, the last work push forward the concept of immersion as its own axis of the user's experience. They insist on its validity in addition to the three core cognitive notions used widely in game studies to describe the user's experience that are cognitive absorption, flow and presence.

Indeed, [31] argues that contrary to immersion, Cognitive Absorption, a term coined by [32] to describe a general attitude and tendency to be involved in software, is not an occasional time limited state of being.

Then, although very similar to flow in its definition, [31] argues that immersion is a much broader term considering how specific the state of flow is as immersive experience may not necessarily meet all the flow criteria. The notion of flow has been defined initially by [33] [34] as a "holistic sensation that people feel when they act with total involvement" in the study of athlete and artist work process. They identified 8 criteria for this state to be achieved:

- clear goals
- direct and immediate feedback
- a balance between challenge and ability
- a sense of personal control over the situation or activity
- a high degree of focus on a limited field of attention
- distorted sense of time
- the merging of action and awareness and a loss of self-consciousness.

[35] argues that flow is applicable to games and the concept has been extensively used in game studies.

Finally, the distinction between immersion and the notion of presence defined earlier is particularly important when dealing with non-realistic games and games with that aren't FPS/TPS. Indeed, [31] argues that a player may not feel present in a game like Tetris [36] while being deeply engaged in the game's tasks.

Engagement has also been studied as a close notion to immersion, although with varying definition. [37] uses it interchangeably with involvement, allowing the existence of passive engagement while[38] bases her definition on an active player that would put its energy to succeed on the account of a given object.

An observation made by many researchers is that immersion and its neighbour concepts are not so well defined [38] [29], present multiple definition or have been used interchangeably in multiple studies. This is the reason why some studies tend to use broad questionnaires that include most of these notions in one form or another.

A plausible cause for that is that most of these concepts are factors extracted to explain axes of variance in data. Thus, the wide range of questionnaires dealing with similar notions of the user experience will promote their own version of these notions [39].

Interestingly, audio in games has been proven to have a major impact on player's immersion [40]. Similar results were found in [41] who separated in their experiment music from intradiegetic audio (sound coming from the game's world, the notion will be expanded upon in 2.4) and found that while diegetic sound increased immersion and flow, the addition of music further improved only immersion and instead deteriorated flow by distracting the players. On another note, [42] argues that in addition to the richness of the game's world (complexity, number of information channels...), consistency of this rich information coming from the game's environment is the other key factor to player's immersion. In this study framework, the hypothesis of visual-sound synchronization having a positive impact on player's immersion could thus be based on the fact that it would provide an increased consistency by adding inter-correlations between information channels.

2.3 Why synchronization?

This raises an important question which is trying to understand to what extent synchronization of visual and sound in art has an effect on the public's experience. An interesting starting point would be to look into dance performances, since as stated in [43], "Synchrony is the central feature of dance". The goal is then to understand what makes dance pleasurable for the dancer but also, more interestingly for this work, for an audience. [44] lists multiple functions of dance that we will expand upon.

First of all, although outside of the scope of this work, there is an evident social component to dance which pushes synchrony to the foreground. Many studies found that synchrony of motion (including dancing) between individuals had various positive effects, both social (affiliation, liking [45]) and individual (increased pain threshold [46]).

Another function listed is the inducing of a flow state for the dancer or the spectator, which is particularly pertinent in the context of video game experiences. According to [47], this could be due to various phenomena of kinesthesic responses (i.e how someone responds to external motion). They identify three categories:

- kinesthesic empathy: the spectator imagine themselves doing the motion and retrieve pleasure similar to the pleasure while dancing ([43]).
- kinesthesic sympathy: the spectator experience admiration for the dancer's effort and virtuosity.
- kinesthesic contagion: emotional reaction to the music and motion, which can includes a synchronization of body functions to the rhythm (breathing, heartbeat...).

Also, games are multisensory experiences. Senses can't be considered autonomous systems who function independently from each other. Human senses are complex, intertwined systems who influence and interact with each other in various ways [48]. Multisensory experiences can increase emotions and impact on the user. For example, [49] showed that haptic effect created more emotionally engaging and stimulating art exhibitions, especially for abstract art. Similarly, [50] showed how adding other senses and interaction in exhibits significantly altered the visitor's experience, and notably helped them enter the flow state.

This relates to the richness of the experience, cited in [42]. However, with this increase in sensory inputs, the importance of "consistency", the second factor of immersion, is intuited to increase. Moreover, as explained in [43], the main factor of music enjoyment is a prediction-anticipation-validation process in the brain. Successful predictions create pleasure while failed predictions bring up short term discomfort that is counteracted with a longer term release of pleasure chemicals (meant to counteract pain in danger situations).

Adding visuals to a music has the effect of creating an information overload [51] which can negatively affect reaction time. This information overload could be an obstacle for the enjoyment of both the visual and the audio of the game as the brain can become unable to properly conduct its prediction process. Synchrony of the two channels could then be an attempt to merge the prediction task and therefore offer an experience of "prediction outcome anticipation" that is richer yet achievable for the human brain. It is interesting to note that the notion of anticipation was already mentioned earlier in this work when talking about the role of music in animation as an "aid for apprehension" of visual motion according to Michel Chion [9]. This hypothesis (although not entirely aligned with the results of [51]) offers another possible explanation of the interest of synchrony, this time as one channel helping the prediction task on the other channel.

2.4 Sound and diegesis

An interesting framework that is pertinent for the subject of this work is the notion of diegesis which is used for sound in film or animation. Intradiegetic sound is supposed to exist inside the narrative framework of the piece (e.g footstep sounds) while extradiegetic sounds are detached from the narrative world and can only be heard by the public and not the character in the movie. As explained in [52], [53] and [54], this dichotomy falls short when dealing with interactive experiences like video games, mostly because the user is now an entity that is both outside and inside the game's diegesis, which requires new categorisation of sound. For example, [53] proposes the addition of transdiegetic sound, which is sound that is in between the two previous categories (either extradiegetic sound closely related to in game events or intradiegetic sound that only exists to give the player information). Similarly, [54] adds two categories: half-diegetic sounds, which are not exclusive to video games and refer to the mixing of sounds both in and out the diegesis (like in The Informer [12] where extradiegetic music and intradiegetic sounds are juxtaposed) and interface sounds which are meant to give signs and feedback to the player.

These notions are interesting for this work because when synchronizing the animations to the soundtrack, the position of the music as outside the game's diegesis is being challenged. There is a layer of complexity to this question that was already raised by a special category of movies: musicals. Indeed, [55] evokes a "diegetic purgatory" to design the existence of a special in between for these movies where music is not acknowledged by the character yet deeply shapes and direct the action and visuals of the movie. Although according to the definitions above, one could argue that synchronization would push the game's music into the transdiegetic category (simply because it does not fit into the traditional diegetic categories), the different cases described by [53] and the fact that this particular situation is not really a result of the addition of player interaction makes it hard to call it a great match. In fact, the pertinence of the notions of diegesis when talking about musicals can be questioned, as these usually portray a surreal depiction of reality.

The comparison with musicals in the case of this work is pertinent. When synchronizing the game visuals to the extradiegetic music, we could argue that instead of moving the music in a different position in regard to the game's diegesis, we instead add a layer of surrealness to the depiction of the game's narrative world and story. This consideration explains why synchronization in non-rhythm games seems to be favored by non-realistic games (Mario 3D World, Mario Kart) that do not rely on realisticness to engage their players.

2.5 Design challenges of synchronisation

Finally, although not at the core of the subject of this work, it is important to understand the challenges that are raised when dealing with sound-visual synchronization in video games. Indeed, with nowadays technologies, synchronisation of visual and sound in film and animation is an aesthetic and production (costly back and forth between visual art and sound) challenge rather than a technical challenge. However, in real-time interactive video game, there is an additional layer of complexity which is the non determinism of the experience. Because the player can interact with the game's world, not every playthrough will be the same.

In fact, for each element of the game's world, there are aspects of it that are influenced by the player, and aspects of it that are not. If we think about an explosive barrel for example, player may be able to change its state by making it explode. On the other hand, the visual of the barrel in each state (size, color, explosion animation...) is usually not impacted by the player's action. This example highlights the added difficulty when dealing with synchronization in games since the player's influence is on timing. The developer can not synchronize the explosion of the barrel to a given timestamp in the game's music because they do not control when (along the music track) the player will make it explode. Moreover, timing of visual events like the barrel explosion are not the only thing that can be subjected to the player's action in a game. Dynamic music is another layer of complexity as the music tracks itself can not in this case be considered predictable.

3 Experiment and method

3.1 Hypothesis

As stated before, the timing of gameplay related visuals is not something that the art direction can easily alter in non rhythmic games without breaking gameplay. For this reason, this work chose to focus on the other animated visuals, namely ambient animated elements (for examples light flickers, wind induced motions...). These elements can be seen as a promising opportunity to balance out the overwhelmingly unilateral subordination relationship of the visual and sound poles of game production (soundscape adapting to the visual landscape of the game) illustrated in 3.1. This could then be done without introducing a huge production cost increase that would happen when creating a loop (which from a production standpoint implies frequent back and forth).

Moreover, as evoked earlier in this paper, adding this type of synchronization could create a layer of surrealness (that can be seen in musicals for example) which is why the focus will be on non-realistic games. This rupture of the separation between extra-diegetic and the game's content can also be seen as an artistic tool that could potentially reinforce some emotions in the player.

Finally, it became apparent that describing the player's experience and its quality is a complex task. In particular, the impact of visual-sound synchronization on aspects of the player's experience has not been experimented a lot on so being too specific in the targeted player's state or emotion would be arbitrary. However, notions like immersion (in all this vagueness, as we explained in 2.2) will be given particular attention.

Considering all of this, the hypothesis of this work is that the synchronization of the animations of the ambient elements onto the music track in non-realistic solo games may have a positive impact on the player experience by increasing immersion.

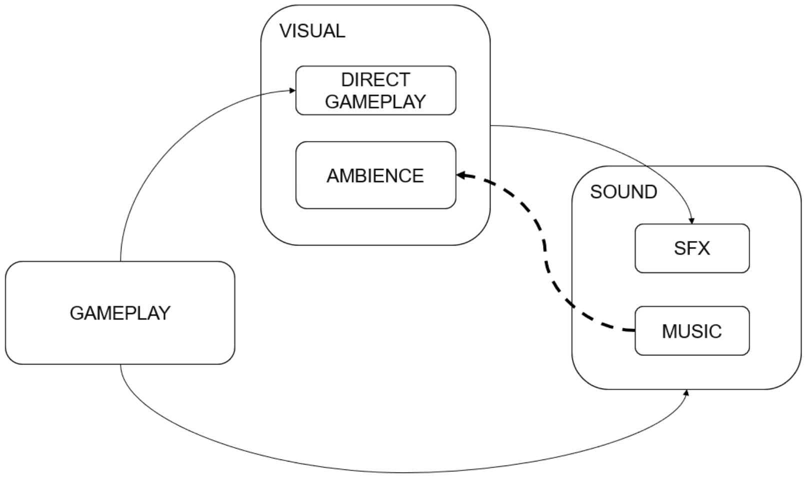


Figure 3.1: Very simplified graph of the common subordination relationships of the different aspect in a game. Dotted arrow represent the proposition of this paper.

3.2 Design

3.2.1 General design of the experiment

To try to answer whether or not synchronization of a stylized game's environment to the musical soundtrack has an impact on the player's immersion, an experiment was designed and conducted. Because the topic at hand can be reduced to whether or not a specific aspect of a game has an impact on particular dimensions of the user's experience, the Experimental Method in its most classical form was deemed to be the appropriate scientific protocol. What this means is that subjects were randomly separated in two groups and were asked to play two almost identical version of a game where the only varying factor was whether or not the ambient effects were synchronized to the music of the game. Using this setup, we then try to measure each subject's immersion and neighbour concepts.

Because the variable that is measured is abstract (has to be extrapolated from other measures) and because the variation between the two groups isn't intuitively expected to be extreme, a significant amount of data needs to be collected, which is why a lot of the decisions concerning the protocol will be influenced by deployment considerations (keeping the duration of the experiment short, platform compatibility...).

3.2.2 Self report measures on player experience

Questionnaires (i.e subjective self report measures) were chosen over interviews or focus groups because they are easier to deploy and adapted for measuring the variation of a specific dimension of the user's experience [56]. Objective trials were not chosen because objective behavioral measures of player immersion are not a well documented subject and no idea of variable to measure seemed interesting or promising enough to warrant spending a significant time setting up a system to collect data on the chosen game.
However, there are many questionnaires that measure part of the player experience [39]. Most prominent examples are:

- the Game Experience Questionnaire (GExpQ) [57] designed to cover multiple aspects of the player's experience and claiming coverage of a range of digital game experiences.
- the Game Engagement Questionnaire (GEngQ) [58] developed to try to measure the consequences of engagement in violent video games.
- the Immersive Experience Questionnaire (IEQ) [59] aimed at measuring the level of immersion of a player.

A big drawback of some of these questionnaires is their length. In the case of this work, finding participants and having them finish the experiment and answer the questionnaire is a significant bottleneck. Therefore, IEQ which presents 31 items, with a lot of similar sounding questions, often not really adapted to the game of our experiment, was not chosen. However, it is to be noted that this is the most scientifically well received questionnaire of the three [60] [39] [61].

Also, although GEngQ has proven to be a valid and reliable questionnaire [60], the measurement is in a unidimensional space (engagement as a single variable) and aims to deal with a precise context (young male subjects playing violent video games).

Finally, GExpQ is the most used in the study of video game experiences although its scientific basic has been highly contested [60] [61]. This is likely due to the fact that this questionnaire answers a need in the field for measures that tackles the multifaceted nature of the game experience. This is especially the case in this work. Indeed, when dealing with a component of a game (in this case visual-audio synchronisation) that have not received much attention from researchers, focusing on measuring a single hyper-precise aspect of the player's experience feels like a shot in the dark.

Because the use of the GExpQ is still widespread in the field, for its multidimensional approach, its modularity and for a lack of a better suited alternative considering the scope of this work, the search for a better questionnaire (which represent a task that could have provided enough material for its own paper) was not deemed a priority. The modules extracted in [57] from GExpQ answers are the following (ranked by their importance in explaining answer variance):

- Sensory and Imaginative Immersion
- Tension
- Competence
- Flow
- Negative Affect
- Positive Affect
- Challenge.

As we can see, these modules relates to the many star concepts of the field of game experience studies that were presented in 2.2. In this small scope work that aims at exploring a rather niche and seldom studied aspect of the player's experience, the multidimensional approach of the GExpQ has been seen as an opportunity for an opening large coverage experiment that could later be refined using more precise and scientifically solid questionnaires or experiments targeting for example specific aspects of the user experience.

3.2.3 Experiment's questionnaire

The experiment questionnaire was composed of three sets of items. First, data quality assessment items, whose purpose was to get an indications of the circumstances in which the game was played. These consisted of 5 items. Among these, "Did you play the game?" and "Did you play the game with the sound on?" were used to discard non relevant data (2 subject's answers were discarded in the experiment). "Which platform did you play the game on?" was asked to detect a posteriori whether or not this introduced a bias and "Which version of the game did you play?" was used as an additional validity check of our control-test separation protocol explained in 3.4.1. Finally, "Did you finish the level?" was added to provide, if needed, a way to ensure all the subjects had played at least a minimal portion (in this case, the entirety) of the game's experience.

Then, a single item about the subject's profile was added ("How often do you play video games?") to get some information on our subject population. In a situation where questionnaire length was not a significant bottleneck, adding more questions about the subject's demographic would have allowed to better detect biases in the participant's sampling (age, status, gaming habits...).

Finally, there were 18 items that aimed at measuring the subject's experience. Just like the GExpQ [57], a 5-point, unipolar intensity-based answering scale was used, with (0) standing for *not at all*, (1) *slightly*, (2) *moderately*, (3) *fairly*, and (4) for *extremely*.

Because the original GExpQ is fairly long (33 items), the In-Game GExpQ [62], which consists of a selection of 14 items from the GExpQ and still measures the same modules, was used. However, some of the items were aimed at complete game experiences that included significant narrative structure for example and therefore didn't really match the game prototype used for this experiment. Therefore, some questions from the In-Game GExpQ were replaced by other questions from the complete GExpQ that targeted the same module and better matched our protocol. For example, "I was interested in the game's story" was replaced by "It was aesthetically pleasing", both items measuring the "Sensory and Imaginative Immersion" module of the GExpQ. It is important to note that this modification may have impaired the scientific validity of the questionnaire but considering its already controversial scientific nature, for the scope of this work, this modification was deemed beneficial to the experiment.

4 additional items, that targeted the specific nature of our problematic were added: "I payed attention to the soundtrack", "I think the soundtrack fitted the experience", "I felt like my action could influence the soundtrack" and "I felt like the soundtrack was connected to the game's environment".

3.3 The game

3.3.1 Design

3.3.1.1 Constraints and focus

The game that would be played as part of the experiment needed to fit a certain number of criteria.

1. **Source code of the game available** This is essential to be able to implement effect synchronization and to produce two builds of the game, with and without effect synchronization.

2. **Multiple ambient animated visual effects** This is necessary to ensure the varying factor between the control and the test setup is significant.

3. **Playable on browser, preferably both on desktop and mobile** This part comes from the deployment considerations. To maximize the number of participants, some frictions need to be avoided (downloading an app would deter most subjects from participating).

4. **Short playtime** The bulk of the experience needed to be accessed quickly (in order not to deter subjects from finishing the experiment). Also, allowing some subjects to play for a much longer time would have introduced a lot of variance as the part of the game that would be used to answer questions would not necessarily be the same (some players would use their experience of the end of the level to answer the questionnaire, while other would be speaking mostly about the beginning). Therefore, a short playtime ensures that most subject experience the same part of the game.

5. **Immersive potential** The aspects of the player experience that we wanted to measure needed to be provided by the game we wanted to play as much as possible. Some games consisting entirely of gray block may not provide immersive experience to their players (depending on the gameplay, the way the narrative frame is introduced to the player...).

6. **High affordance** The users should be able to play the game easily with limited friction.

Constraints 3 and 2 made it really hard to use already existing open source games. Therefore, the choice to create a playable game from scratch for this experiment was made (which automatically satisfies constraint 1).

3.3.1.2 *The game*

The game was made using the game engine Unity [63] for its WebGL support which allows to create browser embedded games and thus satisfy constraint 3.

It is a 2D platforming game where the player incarnate a hermit crab in deep sea waters that must leads a spirit back to its home while avoiding obstacles. Whenever the player get hit, the shell containing the spirit breaks and reappears at its original position, where the player must grab it back. The experience consists in a short level (a few minutes, thus satisfying constraint 4) where the player has to grab the spirit's shell, and bring in to the right end of the level. Figure 3.2 illustrates the visual display of the game.

The gameplay consists of the following actions: walk, jump, grab shell. Knowing the target needed to be both desktop and mobile platforms, control buttons are displayed on the screen that serve both as buttons for mobile players and as control mapping indication for desktop players. This aspect, in addition to dynamic UI prompt indicating the current objectives, level design (direction panels, dynamic will-o'-the-wisps lighting as the player moves in the right direction) and feedbacks (UI shakes, shell break vfx/sfx), helped ensure the constraint 6 about affordance was respected. Considering the scope of this work, affordance in this case does not mean that each object in the game is affordant (as this would require a lot of design work). Instead, UI indications are used maybe more than they normally would to reduce as much as possible the probability of subjects getting stuck (this relates to the complicated topic of onboarding in playtests).

The underwater setting coupled with the ghostly theme created many opportunities for ambient visual effects: waving algaes, swaying will-o'-the-wisps, bubbles, god rays... which amply satisfied constraint 2. Also, the strong theme and the animated visuals and character helped satisfy as much as possible (considering the scope of this work) the constraint 5 about the immersive potential of the game.

3.3.2 Technical process

Apart from the design considerations that aimed at satisfying all the constraints listed in 3.3.1.1, the general notion of synchronizing ambient visual effects to the music of the game had to be concretized and then implemented.

Figure 3.2: Screenshot of the experiment's game.

3.3.2.1 From audio to in game exploitable data

3.3.2.1.1 Data type

The data that can be extracted to help design synchronicity may come from two sources:

- real-time data (like in VJing, that usually uses a frequency spectre computed in real-time): those are especially pertinent when dealing with very dynamic and/or procedural music.

- pre-computed data: simpler to implement, these data can be manually and/or procedurally associated to each music track and then be accessed in game. When using procedural tools, the results are usually more precise than the real-time ones as they have access to the entire music track as input information. For example: prediction of the current bpm is more precise when having access to the upcoming beats and not just the previous ones.

For this experiment, we used a single music loop. Therefore, pre-computated data was the most appropriate approach. The focus was put on beat synchronization (volume or additional data were not used). Intuitively, this was deemed to be the most cost-effective synchronization type in terms of implementation complexity and player experience impact.

3.3.2.1.2 Automatic data extraction

The technical challenge was then to extract the beat information from a given audio file. It is very rare for a piece of music to adhere strictly to a given tempo, therefore, the average BPM is not enough information to properly lead beat synchronization. The Unity add-on [64] for example, gives tools to its users to automatically place events on the beats by analyzing the input audio. The add-on not being in my possession, the software Essentia [65] and its python library were used. More precisely, its implementation of Multi-feature beat tracking [66] was used to extract timestamps of the different beats outside of the game engine and use this data in game for synchronization. The code is available in A.1.

3.3.2.1.3 Analysis

The music used in the game come from [67] and was chosen for its noticeable enough beat structure and how it fitted the game's childish yet calm visuals. The beat tracking algorithm found an average BPM of 81.98BPM with a confidence [68] of 1.55 (confidence above 1.5 means good confidence in the algorithm's prediction). The standard deviation of the BPM was 0.28 BPM. If a timestamp based analysis was not conducted and instead a constant BPM of 81.98 BPM was used to create beat events directly in the game engine, the beat would have been off by 6% in average, with the maximum mismatch being 33% (without consideration for the timing constraints of the engine such as the Update function's *timeDelta*...). In the particular music track used, this is maybe negligible but many pieces have a much more irregular tempo.

3.3.2.2 Visual synchronization methods

3.3.2.2.1 Exploit data in the game engine

Inside the game engine, a coroutine compares the current time with the time when the audio started and progressively moves along through the imported timestamp table. The audio system time (AudioSettings.dspTime) is used instead of the engine's time for better fidelity [69]. In addition to the timestamps, two types of data are exposed for visual effects to use:

- **"Beat progression"**: a continuous value *beat* such that $floor(beat)$ is the last beat's number and $frac(beat)$ is the progression of the current beat (offbeat is when $frac(beat) = 0.5$ for example).

- **Beat events**: events broadcast on the frame a beat (or a specific beat division) happens. This additional data is not “necessary” because redundant with the above data but centralizing the computation of these events is better architecture-wise.

A specific aspect requires additional attention. When stitching two pieces of audio (looping for example), the variable *beat* but also *measure* (where $measure = beat/4ina4/4musicscore$) used to synchronize the environment must remain continuous and accurate. For the game, this was quite simple since the score had a round number of measures but in general, this would require either silence to round up the beat count or fading in and out of a desynchronized motion of similar speed during the transition.

3.3.2.2.2 Shader synchronization

A gobal shader variable can be used to store *beat*. For optimisation purposes (mobile platforms can’t do too much gpu computations) and to prototype visuals faster, more global variables were added: impulse curves or “beat like” curves (almost immediate attack time followed by release without sustain level) synchronized on different time scales for example. All these variables were then used to drive vertex displacement effects or uv distortion texture’s panning speed. An application example would be the algaes that wave according to a sine of *beat* and that also experience a lower frequency underwater impact-like motion (like a strong underwater current) every beat. Other effects that used this were the turtle’s shell, the corals, the anemones and the will-o’-the-wisps.

3.3.2.2.3 Particle systems synchronisation

For particle systems, every n beats, burst emission events that would loop with a given delay were setup, each time with an updated value of delay using the *timestamp* table. Particle lifetime also needed to be updated, since this is the base value that drives most of the particle’s animation. This system was used for the bubbles and god rays.

3.3.2.2.4 Other effects

A few effects were purely driven by scripts and therefore very easy to setup once the dynamic beat data broadcast system was in place. For example, to synchronize the appearance or disappearance of a will-o’-the-wisp along with player advancing in the level (like the geysers in Moana’s "How Far I’ll

Go" extract [70]), one just had to wait for the next beat event to start the animation. Similarly, the vertical wavings of the turtle spriti and the mines were just driven by sines of *beat*.

3.4 Procedure

3.4.1 A/B testing implementation

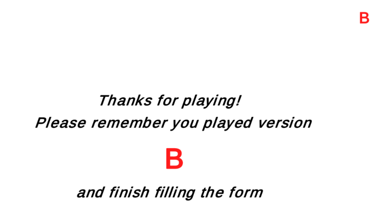
The A/B testing was implemented using a link that would redirect on two separate questionnaires randomly, each containing a link to a different version of the game (with and without synchronization of the visual effects). This way, the data of the control and the test subjects were separated onto two datasets. However, one issue with this setup is that by re-clicking on the link, the subject may end up on a different questionnaire (which may happen if the user tries to go back to the page of the form using a link instead of a return command at the end of the playthrough). This is the reason why a double-check question was added. At the end of the playthrough (and at the top-left of the screen during the entire game), the group's name "A" or "B" is displayed and a question in the form asks for the subject's group ("I don't remember" was left as an option).

In retrospect, these steps were necessary. Indeed, from our 56 participants, 8 reported an answer to the question "Which version of the game did you play?" that did not match with the version associated with the questionnaire they had. Also, 2 participants did not remember the version they played (which is logical because they answered no to at least one of the questions "Did you play the game?" and "Did you finish the level?" and therefore did not have the last warning to remember their version). This means that at approximately 16 participants out of the 56 (29%) clicked on the link again after playing the game and that without adding this explicit mention of the game version played (which admittedly could introduce a bias by breaking the confidentiality about a test/control experiment taking place), about 15% of our data would be wrong.

3.4.2 Test procedure

The test procedure was the following. The subject arrives on the page of one of the two almost identical form (only the link to the game differs). They are prompted to follow the link, play the game and come back to fill the form. When clicking the link, they arrive on a page with a "Start game" button

(a)



(b)

Figure 3.3: In-game screenshots of the beginning (a) and end (b) instructions of a playthrough.

(necessary step to prevent some browsers like Safari to cut-off the game's audio) that launches the WebGL game. The screen displays an indication to play either on a monitor or on a phone in landscape mode, an indication that playing with the sound is necessary and a "Start" button, see Figure 3.3. Upon reaching the end of the level, the subject is thanked, asked to remember the version of the game they played (A or B) and to resume filling the form. The form is separated in two sections. Subjects that answer "No" to the questions "Did you play the game?" or "Did you play with the sound on?" automatically skip the 18 items about player experience.

3.4.3 Recruitment process

To recruit participants, several channels were used. First of all, friends and family that did not know about this work were contacted. Then, online spaces like Facebook groups and subreddits (r/playtesters, r/Unity3D, r/Sample-Size) were queried, each time trying to find a balance between reducing bias and respecting the online space's rules. A general bias that was identified all around was that many participants viewed the experiment material as a game in development (this could be seen in the answers to "Do you have any additional remarks?") which may have impacted some of the answers.

In a larger scale study, it could be interesting to put in place an automatic system to associate a participant's data with their origin (using multiple google form would be complicated to manage if many recruitment channels are used, especially since our setup already requires two forms to work).

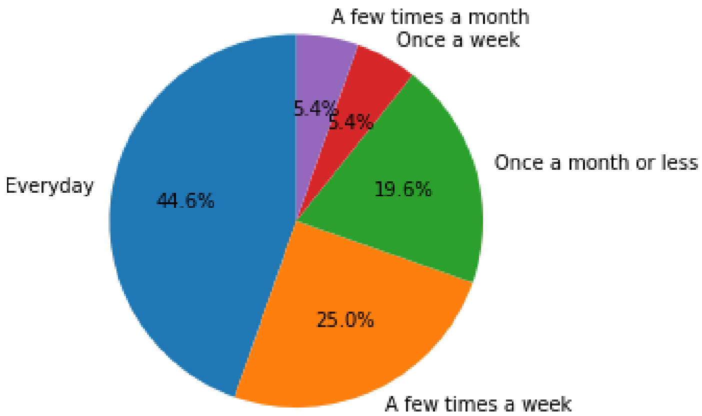


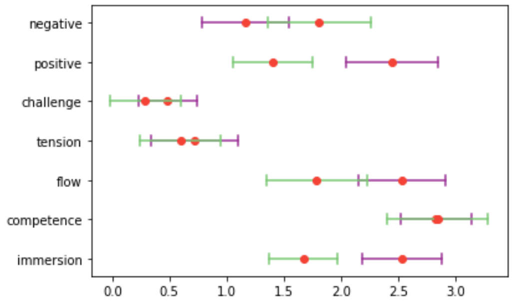
Figure 3.4: Answers of the participants to the question "How often do you play video games ?".

3.5 Participants

56 people filled the form. Among these, 54 input are considered viable (Anwered "Yes" to the questions "Did you play the game?" and "Did you play with the sound on?" and did not answer "I don't remember" to the question "Which version of the game did you play?"). 59% of the viable participants played the synchronized version and 41% the unsynchronized one. Although it may seem uneven, this splitting is in accordance with the fact that we used a random attribution with probability 50% which gives us for 54 samples a confidence interval at 95% of $[36\%, 64\%]$. This is a great illustration of how limited the precision can be when studying datasets with such small sample size.

The playing habits of our participants are illustrated in Figure 3.4. We can see that among our subjects, there is a significant part of daily gamers, namely 45% but also occasional ones (adding a "Never" option would have been preferable) making 20% of the subject group. In retrospect, this item does not give very exploitable information as most statistical studies on player habits measure playtime instead, making it impossible to compare our dataset with global trends. Moreover, the surprisingly high amount of people answering "Once a month or less" raises concern about the framing of the question, as some casual gamers do not always consider parts of their

(a)



(b)

Figure 3.5: Average answer (a) for each item and (b) for each module with 95% confidence interval depending on the platform used to play the game (mobile=purple, desktop=green).

gaming habits (e.g mobile games) into consideration unless the question formulation clearly reminds them to. In general, because of questionnaire length concern, a decision was taken not to include additional items aimed at understanding the demographic. This choice will be discussed in 5.

Finally, we can note that 50% of the participants played the game on mobile (with only one of them reporting significant lags). On the other hand, 48% played on desktop (this does not adds up to 100% since one of the participant was a fervent VR advocate that did not play the experiment material and therefore whose data was discarded). It is likely that the majority of the subjects on mobile would not have completed the experiment if they had to switch to their desktop to accomplish it, which justifies the mobile compatibility considerations when prioritizing data quantity. However, this introduces significant noise in the data by adding an axis of variance for the player's experience. In fact, the answers to the questionnaire differed significantly depending on the platform, see Figure 3.5. As we can see, mobile players had an overall "better" experience, with greater immersion and positive affect, and (with less certainty) higher sense of flow and lower negative affect. This could be partially due to the fact that on desktop, the game is not in full screen.

4 Results

In this analysis of the results of the experiment, we will make the assumption that the grade used for the answers (0=not at all to 4=extremely) is uniform, which allows us to use the simpler algorithms of statistics. This is a substantial assumption, that is most likely not verified with enough precision for some of the following operations to be pertinent, which is why more complex analyses (e.g polytomous Rasch models) would be required if more time was allowed for this work. Code for the data analysis can be found in the annex A.2.

4.1 Results

4.1.1 Confidence intervals

Figures like Figure 3.5 and Figure 4.1 plot the average answer to each item computed over a given sample (e.g participants that played the synchronized version) and the confidence interval for a given probability (see the caption). For example, confidence interval at 95% means that there are 95% chances that the real average (and not the estimated average with our limited sample) falls in the confidence interval. We use T-score instead of Z-score to compute the confidence intervals because the size of the groups are sometimes too small for the central limit theorem approximation to be pertinent.

In our case, there is a tricky added complexity since we are trying to compare measures that were made on separate groups without knowing the correlation between these measures. A common assumption is that they are independent but this would be most likely wrong in this case, since a user that tends to answer positively to the questionnaire (because of their personality, receptiveness to the proposed game experience and/or questionnaire filling tendencies) will have answers above the average in both cases. This case shows positive correlation that would tend to reduce the size of the confidence intervals. On the other hand, we could imagine negative correlations, for example people who would particularly dislike the synchronized version and particularly like the unsychronized version (or vice-versa) compared to the average. We can see that negative correlations seem unlikely to be a significant phenomenon compared to the positive correlations, so in general, when comparing the two groups, we can keep in mind that when the intervals are only slightly intersecting, the real average are likely to be indeed different (this correlation can not be quantified without a different experiment however).

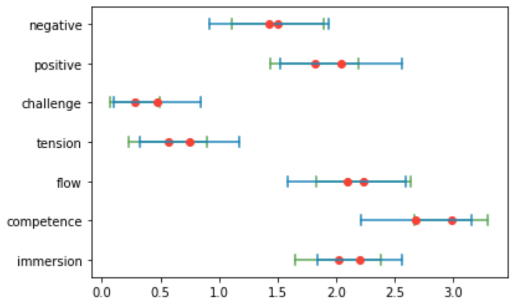
4.1.2 Results

The raw data show little significance between the two study groups. Indeed, as we can see in Figure 4.1, the only significant difference between the two groups is that the test group (synchronized) answered lower on the following items "I feel like my actions could influence the soundtrack", "I payed attention to the soundtrack" and higher on the following item "I was fast at reaching the game's target" compared to the control group. Even for these items, where difference in the average answer has been proven regardless of the sampling noise, the effect is very marginal.

When it comes to the modules of the GExpQ [57], no significant difference between the control and test group can be extracted. In fact, if we disregard confidence intervals and directly look at the average estimates, the results are rather against the hypothesis as immersion, positive affect, challenge and tension are felt a little bit more when the animations are not synchronized while flow and competence are slightly stronger in the participants that played the synchronized version.

In conclusion, the experiment did not manage to measure any effect that would confirm our hypothesis.

(a)



(b)

Figure 4.1: Average answer (a) for each item and (b) for each module with 68% confidence interval (unsynchronized/control=green, synchronized/test=purple).

5 Discussion

As we could see, the experiment's results tend to infirm the hypothesis. To evaluate the quality of the conclusions that were drawn, it is important to reflect on the experiment, its biases and its pertinence in regard to this paper's problematic.

5.1 Experiment bias

As showed in 3.5, the platform used to play had a significant impact on the player's experience. In this case, the splitting between platform was independent on the control-test splitting, so it is unlikely to have introduced any bias but this is a great example to illustrate how small aspects of the experiment can significantly alter the results.

First of all, because we relied on a participant's answer to determine whether or not they belonged to the test or control group ("Which version of the game did you play?"), this represents a potential source of error. Moreover, revealing the existence of several versions of the game with this question is an notable alteration of the classical experimental setup, which could have introduced bias in the sense that it may have made subjects on the lookout for the specificity of their version.

In addition, multiple potential biases can be identified. First of all, the hypothesis did not specify a specific target demographic. Therefore, if the participant pool was supposed to represent an average human sample in the most universal way possible, it was very biased since the subjects were social media users, in specific online groups. Although specifying a target for such a general hypothesis is not common practice in this research field, in retrospect, it would be a good practice to add items to better understand a posteriori the subject base of the tests. This would allow us to give a better context to the potential conclusion of the experiment and the generalizability of the results.

Moreover, depending on the channel used to recruit participants, the way the experiment was introduced differed. This is especially obvious in some of the answers to the "Do you have additional remarks?" where some respondents made reference to the experiment's game as if it was a real project in development (more likely to happen with r/Unity3D or r/playtesters users for example). This additional variance and set of assumptions among participants may have affected the results.

Another potential bias is the speed of the visual effects. Indeed, when toggling the synchronization on/off, some effects won't have the exact same speed (or else this would mean that they would be mostly synchronized to the music in both versions and it would therefore defeat the purpose of the experiment). There was an attempt to input similar speeds and therefore it may not have had a huge impact but it's something to be aware of, in addition to the fact that synchronized animations have less degrees of freedom for tweaking the global motion of the scene.

Finally, it is important to keep in mind that, while they are overwhelmingly common practice in this research field, questionnaires like the one we used are subjective self-reported data and therefore introduce their fair share of biases and uncertainties. Indeed, the way participants answer the questionnaire, variation in the understanding of the questions, wording and the scale used for the answers [39] can all alter the results. On a larger scope work, it would be interesting to analyse the Likert scale we used for the answers (0 to 4) (for example using polytomous Rasch models). Indeed, there are many debates about how to process this kind of data [71], and proper analysis would therefore require additional literature research.

5.2 Experiment design

Apart from the inherent biases of the experiment that can influence the quality of the extracted results, the pertinence of the experiment and its limitations can also be questioned.

As explained in 3.2.2, the questionnaire we used (for specific reasons such as its multidimensionnality and its broad outlook on player experience) has a questionable scientific validity. First of all, when it comes to the GExpQ, the reproductability of the identified modules has been questioned [61], which means that the module scores ("challenge", "positive affect"...) we used in this paper may not be reliable. Cronbach's alpha [72] was computed for our data as seen in Table 5.2. Cronbach's alpha is a tool from classical test theory that aims at measuring the reliability of a questionnaire's measures. We are in a quite particular case since only two items measure each module and we

Sensory and imaginative immersion	-1.03
Competence	-0.33
Flow	-0.79
Tension	-0.54
Challenge	-0.28
Negative affect	-0.25
Positive affect	-0.56

Table 5.1: Cronbach's alpha for the modules of the GExpQ questionnaire on our data.

can not ensure satisfaction of the prerequisite for using this tool but this gives us a rough assessment of the quality of our measure (and an indication of useful methodology if this work were to have a bigger scope). Values should be close to 1, so we can already see that our measures are not very reliable, especially when it comes to immersion and flow. This is not very surprising as we added a significant disturbance (using a shorter version by arbitrarily choosing questions from each modules) to a questionnaire that already fell short on these reliability metrics.

Also, many interrogations emerged while designing a game that would be pertinent for the experiment. Many choices felt arbitrary or motivated by deployment consideration more than user experience considerations. The motivations were to produce an experience that had the potential to be immersive with a low time budget, and the choice to prioritize the art consistency (motivated by [42]) above the gameplay complexity and quality is arguable. Indeed, answers to the "Do you have any additional remarks?" item showed that some participant were confused by the lack of depth of the gameplay. Also, the camera of the game (2D side scroller) was chosen out of concern for the WebGL and mobile performance. However, this is not necessarily the type of game where immersion is the easiest to measure. According to [29], using FPS games can simplify research made on immersion and neighbour notions (flow, presence) by removing potential frictions when it comes to the identification with the character. For example, in the study [73], the game Half-Life 2 [74] was ranked highest in all dimensions (sensory, challenge-based and imaginative immersion) of their gameplay experience model.

Then, the choice of the music is another decision that probably had a determining impact on the result but did not feel grounded in research and ended up being a result of art direction consistency and licensing considerations. The rhythm is very repetitive and simple to grasp which may seem like a great choice for this experiment but this is debatable. [75] found an inverted U-shaped correlation between amount of syncopation and experienced pleasure in listeners of groove music. This relates to [43]. This explains why there is a sweet spot in the complexity or unpredictability of the music when trying to maximize pleasure since the brain would tend to seek this failed prediction situation in a moderate amount. Therefore, if we consider the hypotheses that the results of [75] can be generalized to music rhythmic complexity and that synchronization simplifies the prediction task, synchronization would then make the player experience worse if the music is already simple rhythmically (which is the case of the music track used in the experiment). This hypothesis is illustrated in Figure 5.1 which shows how this interpretation of synchronization would predict contrary effects on pleasure depending on the music's complexity.

This notion of sweet spot could also apply to the amount of synchronization. Excessive synchronization is something that was criticized by some mickeymousing detractors, as evoked in 2.1. Intuitively, it would seem pertinent introduce an axis measuring the "amount" of synchronization when studying synchronization's impact on the user's experience. This amount would then have to be properly defined. In fact, one can think of multiple ways to gradually synchronize visuals to the sound. The number of synchronized elements is one axis that can be leveraged (for example in Mario 3D World [27], only some environmental objects like the flowers are synchronized to the music) but one could also think of creating random functions to drive visual effects that would occasionally synchronize the effect to the sound with exposed control on the probability.

In addition, in the experiment's game, synchronization of the visual to the sound was used as a very systematic process with limited art direction. First of all, only the synchronization to the beat was used (no volume or pitch data). Then, every effect was triggered at regular fixed intervals aligned with the music subdivisions. This is a very rigid framework that does not empower the artists, which is not an inherent property of synchronization (we could see in 3.3.2.2 that the technical tools are in fact very flexible). Therefore with more time allocated, more subtle and interesting use of these tools could be made.

Also, more generally, the paradigm for this study was to prioritize quantity of data over their quality. This choice was motivated by the knowledge that we were trying to measure a very marginal effect. This motivated the

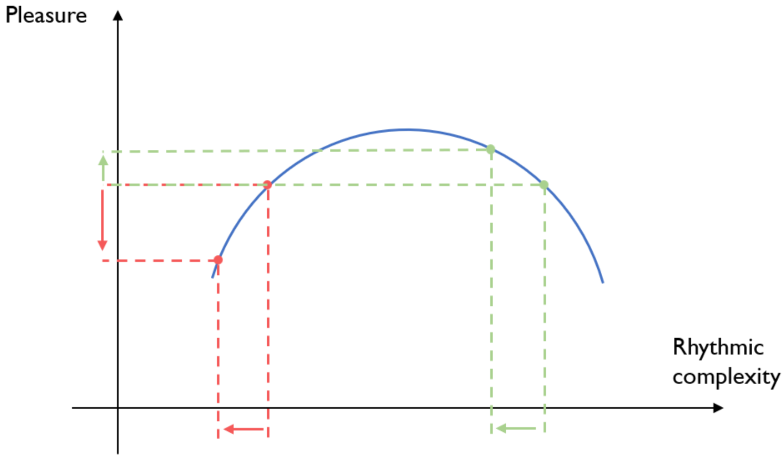


Figure 5.1: Illustration of a potential effect of synchronization (represented as arrows) under the hypotheses of generalisation of the U-Shape curve of [75] and simplification effect of visual-sound synchronization on a music perceived rhythmical complexity.

choice to deploy autonomous self-reported playtests, to propose both mobile and desktop controls and to intensely limit the size of the questionnaire. These two first choices introduced significant amount of noise in the data that may have counterbalanced the benefits of a larger participant pool. On the other hand, the last choice made it hard to extract exploitable information about the player's experience and made the module from the GExpQ even less reliable than they already were. However, questionnaire size is a real concern and can affect answer quality. This is especially limiting when we are tackling a subject that has not been extensively documented and therefore when we can not yet pinpoint a specific dimension of the user experience to measure. For this reason, when trying to first approach a topic, qualitative studies with less participants may be a preferable entry point. These could then be used as groundwork to design a larger scale quantitative experiment. All in all, we can safely conclude that if this hypothesis was to be tackled again in a larger scope work, this experiment protocol would not be the perfect approach as it does not seem to efficiently tackle such subtle effects.

One of the key difficulty that arose from designing an experiment for our hypothesis is the question of measuring marginal variations of complex abstract notions inherent to the player experience such as immersion or flow. The choice that was made was to use a complete (gameplay, sound, visual) game experience and a game experience questionnaire as protocol for the measures. However, the change between control and test groups could be seen as "just a layer of polish".

If the objective was to measure feelings that did not require some form of interaction, this experimental protocol would usually not be chosen. For example, to try and understand user's feeling about color palettes, a user experience researcher would most likely not prescribe a game playtest followed by a user experience questionnaire but rather use still images or videos. This was the case of [3], who conducted their studies on voice desynchronization by using videos instead of playtests and then argued about the generalizability of their findings. This was easy to setup in their case because the emotions they wanted to measure (fear...) are not as intricately linked to the interactivity like for say immersion or presence. If they used instead entire interactive game sequences, they may have had introduced too much noise for any measurable difference to become significant. Therefore, this raises the particularly tricky question: "how to measure marginal effects on immersion, presence, flow and other notions that are linked to interaction?". Using genres like FPS [29] and reducing gameplay noise (simple walking simulator) could be interesting directions.

Finally, it is important to note that the study of visual-sound synchronization effects in a vacuum rather than as an element of the art direction might fail to provide any meaningful results. What this means is that a sprite or a visual effect's impact on the player is not absolute and heavily depends on the surrounding assets and the general art direction of the game. An obvious example is color, whose perception is extremely relative. This might thus also apply to synchronization of the ambient animations to the music (which explains why examples are found mostly in cartoonish games) and therefore highlights a limitation of using a single video game experiment to answer this paper's problematic.

5.3 Improvement proposition

For this reason, ideas for an improved experimental protocol (that does not account for the difficulties to find enough participants) are the following.

The experiment would still be composed of a playable experience and a questionnaire. The game would be composed of multiple scene with various ambiances, art direction and musical tracks (with varying rhythmical complexity). Unless the problematic is targeted at a specific game's genre, player would see in first person camera, to try to maximize the measurably of metrics like immersion or presence. Although the star medium when it comes to immersion, the use of VR in this case may not be advised as the notions of music and diegesis in VR games have their own particularities [52] and therefore findings may not be generalizable to other game genres.

Longer questionnaires with scientific credibility like the IEQ [59] would be used, to which we could then add additional questions to understand the demographic we surveyed to give better context to potential findings. What could be pertinent would be to test a large version of the questionnaire on a small sample of people to be able to remove unnecessary or redundant question with a data analysis of these preliminary results. Mixing several questionnaire is a practice used in several studies plus items specific to sound synchronization like the one used in our experiment could be added. We could then craft a more condensed version that could be deployed on a larger subject base.

6 Conclusion

Synchronization of visual and music in cinema is an artistic practice that went from being a key component of the rise of animation films to a more niche practice used sporadically, mainly in the form of musical-like sequences. Video games could offer a privileged ground for this practice because they offer a multiplicity of action and animated dynamics. Player, Non Playable Characters and the environment all are infused with their own more-or-less independent motion and it is often up to the player to choose which one to focus on (especially when the camera is dynamic). When it comes to synchronization of visuals to the music, it therefore does not have to be restricted to the specific genre of rhythm or music based games and can leverage the environment's animations to avoid impacting the gameplay.

This paper aimed at understanding whether or not synchronization of the ambient animations to the game's music track would have an impact on the player's experience. The analysis of the experiment's participants' answers to a customized version of the GExpQ questionnaire after playing a game with or without synchronization of environmental visual effects to the game's music did not show any significant difference. It becomes especially obvious when comparing the results to other finding such as user experience between mobile and desktop players which had much more glaring differences. It is important to note that although this work did not show any significant measurable impact of the synchronization on notions such as immersion, flow or positive/negative affect, this does not imply that synchronization has no impact in general. Indeed, because of the restricted nature of the experiment (only one game was tested) and multiple other considerations that have been evoked in 5 (both biases of the experiment and more general remarks about its pertinence), the generalizability of these results does not seem arguable.

The process of making a game with synchronized visual effects was also an opportunity to better understand the development cost of this feature. Conclusions made on this topic is that with a minimal effort required to develop appropriate tools, synchronization ended up being quite easy to implement and integrate in a game architecture and did not raise particular concerns about scalability or modularity of the feature. In fact, from a purely production and collaboration standpoint, some visual effects can be exposed as additional "musical instruments" for the artists to exploit with a very reasonable cost in development.

In conclusion, this work highlighted the pluridisciplinarity and complexity of a not so well researched aspect of the user experience. For this reason, this work presents itself more as a kick in the anthill, that unveiled a variety of related issues and offered new ideas as to how to experiment on this topic rather than as a definite answer to its problematic.

Bibliography

[1] Hoyoverse. *Genshin Impact*. [Android, iOS, Windows, PS4, PS5]. China, 2020.

[2] Ryan Clark. *Game Developer Game Design Deep Dive: Finding the beat in Crypt of the NecroDancer*. 2014. URL: https://www.gamedeveloper.com/audio/game-design-deep-dive-finding-the-beat-in-i-crypt-of-the-necrodancer-i- (visited on 10/15/2022).

[3] Angela Tinwell, Mark Grimshaw-Aagaard, and Andrew Williams. "Uncanny behaviour in survival horror games". In: *Games Computing and Creative Technologies: Journal Articles (Peer-Reviewed)* 2 (May 2010). DOI: 10.1386/jgvw.2.1.3_1.

[4] Elliot D. Freeman et al. "Sight and sound out of synch: Fragmentation and renormalisation of audiovisual integration and subjective timing". In: *Cortex* 49.10 (2013), pp. 2875–2887. ISSN: 0010-9452. DOI: https://doi.org/10.1016/j.cortex.2013.03.006. URL: https://www.sciencedirect.com/science/article/pii/S0010945213000890.

[5] Kendra Preston Leonard. *Taking a Cue: Accompanying Early Film*. URL: https://newmusicusa.org/nmbx/taking-a-cue-accompanying-early-film/ (visited on 10/25/2022).

[6] *Don Juan*. 1926.

[7] *Ko-Ko Song Car-Tunes*. 1924.

[8] *Steamboat Willie*. 1928.

[9] Laura Lazarescu-Thois. "From Sync to Surround: Walt Disney and its Contribution to the Aesthetics of Music in Animation". In: *The New Soundtrack* 8 (Mar. 2018), pp. 61–72. DOI: 10.3366/sound.2018.0117.

[10] Daniel Goldmark. *Tunes for 'Toons: Music and the Hollywood Cartoon*. 1st ed. University of California Press, 2005. ISBN: 9780520236172.

[11] Carlos Baena. *Animation as Visual Music*. URL: https://www.animationmentor.com/blog/animation-as-visual-music/ (visited on 10/20/2022).

[12] *The Informer*. 1935.

[13] *Alexandre Nevski*. 1938.

[14] Nicklas Nygren. *UUrnog UUrnlimited*. [Windows]. 2017.

[15] Peter Wegele. *Max Steiner: Composing, Casablanca, and the Golden Age of Film Music*. URowman & Littlefield, 2014. ISBN: 1442231130.

[16] Robert Spadoni. *Uncanny Bodies: The Coming of Sound Film and the Origins of the Horror Genre*. University of California Press, 2007. ISBN: 9780520251229.

[17] *Dracula*. 1931.

[18] *Frankenstein*. 1931.

[19] BioWare. *Mass Effect*. [Windows, Xbox360, Playstation 3]. Canada, 2007.

[20] Jonathan Cooper. *On Mass Effect Andromeda's Dialogue Animation*. URL: https://twitter.com/i/events/845012187231895552?ref_src=twsrc%5C%5Etfw (visited on 10/22/2022).

[21] Stephen DDungu. *Sword Of Symphony*. [In development].

[22] Dean Herbert. *osu!* [Windows, OS X, Linux]. Australia, 2007.

[23] Brace Yourself Games. *Crypt of the NecroDancer*. [Windows, OS X, Linux, Playstation 4, Playstation Vita, Xbox One, Nintendo Switch]. Canada, 2014.

[24] Timothy Sanders and Paul A. Cairns. "Time perception, immersion and music in videogames". In: *BCS HCI*. 2010.

[25] John A. Sloboda. "Music Structure and Emotional Response: Some Empirical Findings". In: *Psychology of Music* 19 (1991), pp. 110–120.

[26] Nintendo. *Mario Kart 8*. [Wii U]. Japan, 2014.

[27] Nintendo. *Mario 3d World*. [Wii U]. Japan, 2013.

[28] Emily Brown and Paul A. Cairns. "A grounded investigation of game immersion". In: *CHI EA '04*. 2004.

[29] Lennart E. Nacke and Craig A. Lindley. "Affective Ludology, Flow and Immersion in a First- Person Shooter: Measurement of Player Experience". In: (2010). DOI: 10.48550/ARXIV.1004.0248. URL: https://arxiv.org/abs/1004.0248.

[30] Matthew Lombard and Theresa Ditton. "At the Heart of It All: The Concept of Presence". In: *J. Comput. Mediat. Commun.* 3 (1997).

[31] Charlene Jennett et al. “Measuring and Defining the Experience of the Immersion in Games”. In: *International Journal of Human-Computer Studies* 66 (Sept. 2008), pp. 641–661. DOI: 10.1016/j.ijhcs.2008.04.004.

[32] Ritu Agarwal and Elena Karahanna. “Time Flies When You’re Having Fun: Cognitive Absorption and Beliefs about Information Technology Usage”. In: *MIS Quarterly* 24 (Dec. 2000), pp. 665–694. DOI: 10.2307/3250951.

[33] Mihály Csíkszentmihályi. *Beyond boredom and anxiety*. San Francisco: Jossey-Bass Publishers, 1975.

[34] Mihály Csíkszentmihályi. *Flow: The Psychology of Optimal Experience*. New York: HarperPerennial, 1975.

[35] Benjamin Cowley et al. “Toward an understanding of flow in video games”. In: *Comput. Entertain.* 6 (July 2008), pp. 1–27. DOI: 10.1145/1371216.1371223.

[36] Alekseï Pajitnov. *Tetris*. Soviet Union, 1984.

[37] Kelly Johnston. “Engagement and Immersion in Digital Play: Supporting Young Children’s Digital Wellbeing”. In: *International Journal of Environmental Research and Public Health* 18.19 (2021). ISSN: 1660-4601. DOI: 10.3390/ijerph181910179. URL: https://www.mdpi.com/1660-4601/18/19/10179.

[38] Delphine Soriano. “Outils d’évaluation du design de l’avatar dans l’engagement du joueur”. Theses. Conservatoire national des arts et metiers - CNAM, Dec. 2016. URL: https://tel.archives-ouvertes.fr/tel-01968065.

[39] A. Imran Nordin and Alena Denisova. “Too Many Questionnaires: Measuring Player Experience Whilst Playing Digital Games”. In: Oct. 2014.

[40] Jiulin Fu. “The Influence of Background Music of Video Games on Immersion”. In: *Journal of Psychology & Psychotherapy* 05 (Jan. 2015). DOI: 10.4172/2161-0487.1000191.

[41] Mark Grimshaw-Aagaard. “Sound and Immersion in the First-Person Shooter”. In: *Games Computing and Creative Technologies: Conference Papers (Peer-Reviewed)* (Jan. 2007).

[42] Jamie Madigan. *The Psychology of Immersion in Video Games*. 2010. URL: https://www.psychologyofgames.com/2010/07/the-psychology-of-immersion-in-video-games/ (visited on 10/18/2022).

[43] Eloise A. Stark, Peter Vuust, and Morten L. Kringelbach. "Chapter 7 - Music, dance, and other art forms: New insights into the links between hedonia (pleasure) and eudaimonia (well-being)". In: *The Arts and The Brain*. Ed. by Julia F. Christensen and Antoni Gomila. Vol. 237. Progress in Brain Research. Elsevier, 2018, pp. 129–152. DOI: https://doi.org/10.1016/bs.pbr.2018.03.019. URL: https://www.sciencedirect.com/science/article/pii/S0079612318300190.

[44] Julia Christensen, Camilo Cela-Conde, and Antoni Gomila. "Not all about sex: neural and biobehavioral functions of human dance: Neural and biobehavioral functions of human dance". In: *Annals of the New York Academy of Sciences* 1400 (July 2017), pp. 8–32. DOI: 10.1111/nyas.13420.

[45] Michael Hove and Jane Risen. "It's All in the Timing: Interpersonal Synchrony Increases Affiliation". In: *Social Cognition, v.27, 949-961 (2009)* 27 (Dec. 2009). DOI: 10.1521/soco.2009.27.6.949.

[46] Bronwyn Tarr et al. "Synchrony and exertion during dance independently raise pain threshold and encourage social bonding". In: *Biology letters* 11 (Oct. 2015). DOI: 10.1098/rsbl.2015.0767.

[47] Matthew Reason and Dee Reynolds. "Kinesthesia, Empathy, and Related Pleasures: An Inquiry into Audience Experiences of Watching Dance". In: *Dance Research Journal* 42.2 (2010), pp. 49–75. DOI: 10.1017/S0149767700001030.

[48] Matthew Fulkerson. "Explaining Multisensory Experience". In: July 2014, pp. 365–373. ISBN: 978-94-007-6000-4. DOI: 10.1007/978-94-007-6001-1_25.

[49] Chi Thanh Vi et al. "Not just seeing, but also feeling art: Mid-air haptic experiences integrated in a multisensory art exhibition". In: *International Journal of Human-Computer Studies* 108 (2017), pp. 1–14. ISSN: 1071-5819. DOI: https://doi.org/10.1016/j.ijhcs.2017.06.004. URL: https://www.sciencedirect.com/science/article/pii/S1071581917300988.

[50] Mark L. Harvey et al. “The Influence of Museum Exhibit Design on Immersion and Psychological Flow”. In: *Environment and Behavior* 30.5 (1998), pp. 601–627. DOI: 10.1177/001391659803000502. eprint: https://doi.org/10.1177/001391659803000502. URL: https://doi.org/10.1177/001391659803000502.

[51] Kyung Myun Lee et al. “Dance and Music in "Gangnam Style": How Dance Observation Affects Meter Perception”. In: *PloS one* 10 (Aug. 2015), e0134725. DOI: 10.1371/journal.pone.0134725.

[52] Isak Bosman. “The Role of Audio in the Diegesis of Virtual Reality Games”. In: Apr. 2022. DOI: 10.1145/3491101.3503810.

[53] Kristine Jrgensen. “On transdiegetic sounds in computer games”. In: *Northern Lights: Film and Media Studies Yearbook* 5 (Sept. 2007), pp. 105–117. DOI: 10.1386/nl.5.1.105_1.

[54] Isabella van Elferen. “¡Un Forastero! Issues of Virtuality and Diegesis in Videogame Music”. In: *Music and the Moving Image* 4.2 (2011), pp. 30–39. ISSN: 21678464, 19407610. URL: http://www.jstor.org/stable/10.5406/musimoviimag.4.2.0030 (visited on 10/23/2022).

[55] Maclen Johnson and Jacob Maher. *In honor of “A Star Is Born” (2018), an analysis of diegetic musical movies*. 2018. URL: http://www.drurymirror.org/2018/10/12/in-honor-of-a-star-is-born-an-analysis-of-diegetic-musical-movies/ (visited on 10/23/2022).

[56] Anne Adams and Anna Cox. “Questionnaires, in-depth interviews and focus groups”. In: *Research Methods for Human-Computer Interaction* (Jan. 2008).

[57] Karolien Poels, de Yaw Yvonne Kort, and Wa Wijnand IJsselsteijn. “D3.3 : Game Experience Questionnaire:development of a self-report measure to assess the psychological impact of digital games”. In: 2007.

[58] Jeanne Brockmyer et al. “The development of the Game Engagement Questionnaire: A measure of engagement in video game-playing”. In: *Journal of Experimental Social Psychology* 45 (July 2009), pp. 624–634. DOI: 10.1016/j.jesp.2009.02.016.

[59] Charlene Jennett et al. “Measuring and Defining the Experience of the Immersion in Games”. In: *International Journal of Human-Computer Studies* 66 (Sept. 2008), pp. 641–661. DOI: 10.1016/j.ijhcs.2008.04.004.

[60] Kent Norman. "GEQ (Game Engagement/Experience Questionnaire): A Review of Two Papers". In: *Interacting with Computers* 25 (June 2013), pp. 278–283. DOI: 10.1093/iwc/iwt009.

[61] Lai-Chong Law, Florian Brühlmann, and Elisa Mekler. "Systematic Review and Validation of the Game Experience Questionnaire (GEQ) – Implications for Citation and Reporting Practice". In: Oct. 2018. DOI: 10.1145/3242671.3242683.

[62] Wa Wijnand IJsselsteijn, de Yaw Yvonne Kort, and Karolien Poels. "The Game Experience Questionnaire". In: 2013.

[63] Unity Technologies. *Unity®*. Version 2021.3.9f1. Oct. 16, 2022. URL: https://unity.com/fr.

[64] Sonic Bloom. *Koreographer Unity Add-on*. Version 1.6.1. Oct. 15, 2022. URL: https://www.koreographer.com/koreographer.

[65] Music Technology Group (UPF). *Essentia*. Version 2.0.1. Oct. 15, 2022. URL: https://essentia.upf.edu/licensing_information.html.

[66] Jose Zapata, Matthew Davies, and Emilia Gómez. "Multi-Feature Beat Tracking". In: *Audio, Speech, and Language Processing, IEEE/ACM Transactions on* 22 (Apr. 2014), pp. 816–825. DOI: 10.1109/TASLP.2014.2305252.

[67] Zapsplat. *Zapsplat.com*. URL: https://www.zapsplat.com/ (visited on 10/16/2022).

[68] Jose Zapata et al. "Assigning a confidence threshold on automatic beat annotation in large datasets". In: *Proceedings of the 13th International Society for Music Information Retrieval Conference, ISMIR 2012* (Jan. 2012).

[69] Chistian Floisand. *Beat Synchronisation in Unity*. URL: https://christianfloisand.wordpress.com/2014/01/23/beat-synchronization-in-unity/ (visited on 10/15/2022).

[70] *Moana*. 2016.

[71] Fabio Antonialli, Luiz Antonialli, and Renan Antonialli. "Uses and Abuses of the Likert Scale: Bibliometric Study in the Proceedings of EnANPAD From 2010 to 2015". In: 22 (Jan. 2018), pp. 1–19. DOI: 10.21714/2179-8834/2017v22n4p1-19.

[72] Lee Joseph Cronbach. "Coefficient alpha and the internal structure of tests". In: *Psychometrika* 16 (1951), pp. 297–334.

[73] Laura Ermi and Frans Mäyrä. “Fundamental Components of the Gameplay Experience: Analysing Immersion.” In: Jan. 2005.

[74] Valve Corporation. *Half Life 2*. [Windows, OS X, Linux, Xbox, Xbox360, Playstation 3]. United States, 2004.

[75] Maria A G Witek et al. “Syncopation, body-movement and pleasure in groove music”. English. In: *P L o S One* 9.4 (2014), e94446. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0094446.

A Appendix

A.1 Extracting beat timestamp: python code

`https://colab.research.google.com/drive/1CSABfA9aMS6sBSsnrpCTz6VIg6y0cenL?usp=sharing`

A.2 Questionnaire result analysis: python code

`https://colab.research.google.com/drive/1XbTMkt-rB_1XaYuoX7FmHrjAjK4uiiw0?usp=sharing`

A.3 Questionnaire items

- Did you play the game? (the link is in the questionnaire description)
  (Yes; No)
- Did you play with the sound on?
  (Yes; No; I'm not sure)
- Which platform did you play the game on?
  (Mobile; Desktop; Other)
- Which version of the game did you play?
  (Version A; Version B; I don't remember)
- Did you finish the level?
  (Yes; No; I don't know)
- It was aesthetically pleasing
  (0: not at all; 1: slightly; 2: moderately; 3: fairly; 4: extremely)
- I payed attention to the soundtrack
  (0: not at all; 1: slightly; 2: moderately; 3: fairly; 4: extremely)
- I felt bored
  (0: not at all; 1: slightly; 2: moderately; 3: fairly; 4: extremely)

- It felt like a rich experience
  (0: not at all; 1: slightly; 2: moderately; 3: fairly; 4: extremely)
- I felt pressured
  (0: not at all; 1: slightly; 2: moderately; 3: fairly; 4: extremely)
- I forgot everything around me
  (0: not at all; 1: slightly; 2: moderately; 3: fairly; 4: extremely)
- I think the soundtrack fitted the experience
  (0: not at all; 1: slightly; 2: moderately; 3: fairly; 4: extremely)
- I was good at it
  (0: not at all; 1: slightly; 2: moderately; 3: fairly; 4: extremely)
- I enjoyed it
  (0: not at all; 1: slightly; 2: moderately; 3: fairly; 4: extremely)
- I felt frustrated
  (0: not at all; 1: slightly; 2: moderately; 3: fairly; 4: extremely)
- I found it tiresome
  (0: not at all; 1: slightly; 2: moderately; 3: fairly; 4: extremely)
- I felt like my actions could influence the soundtrack
  (0: not at all; 1: slightly; 2: moderately; 3: fairly; 4: extremely)
- I felt tense
  (0: not at all; 1: slightly; 2: moderately; 3: fairly; 4: extremely)
- I was fast at reaching the game's targets
  (0: not at all; 1: slightly; 2: moderately; 3: fairly; 4: extremely)
- I was fully occupied with the game
  (0: not at all; 1: slightly; 2: moderately; 3: fairly; 4: extremely)
- I felt like the soundtrack was connected to the game's environment
  (0: not at all; 1: slightly; 2: moderately; 3: fairly; 4: extremely)
- I thought it was hard
  (0: not at all; 1: slightly; 2: moderately; 3: fairly; 4: extremely)
- I thought it was fun
  (0: not at all; 1: slightly; 2: moderately; 3: fairly; 4: extremely)
- Do you have any additional remarks ?