ASSESSING THE EFFECT OF ADAPTIVE MUSIC ON PLAYER NAVIGATION IN VIRTUAL ENVIRONMENTS

Manuel López Ibáñez

Department of Software Engineering and Artificial Intelligence
Complutense University of Madrid
Madrid, Spain
manuel.lopez.ibanez@ucm.es

Nahum Álvarez

Artificial Intelligence Research Center
National Institute of Advanced Industrial Science and Technology
Tokyo, Japan
nahum.alvarez@aist.go.jp

Federico Peinado

Department of Software Engineering and Artificial Intelligence
Complutense University of Madrid
Madrid, Spain
email@federicopeinado.com

ABSTRACT

Through this research, we develop a study aiming to explore how adaptive music can help in guiding players across virtual environments. A video game consisting of a virtual 3D labyrinth was built, and two groups of subjects played through it, having the goal of retrieving a series of objects in as short a time as possible. Each group played a different version of the prototype in terms of audio: one had the ability to state their preferences by choosing several musical attributes, which would influence the actual spatialised music they listened to during *gameplay*; the other group played a version of the prototype with a default, non-adaptive, but also spatialised soundtrack. Time elapsed while completing the task was measured as a way to test user performance. Results show a statistically significant correlation between player performance and the inclusion of a soundtrack adapted to each user. We conclude that there is an absence of a firm musical criteria when making sounds be prominent and easy to track for users, and that an adaptive system like the one we propose proves useful and effective when dealing with a complex user base.

1. INTRODUCTION

Most video game design challenges are related to the scope of possible player decisions. Current-generation open-world games offer an enormous variety of places to go and things to do, which makes designing specific player behaviour a daunting task. The problem of guiding a player through a big and complex virtual environment is frequently solved by adding extradiegetic information to the graphical user interface (GUI), thus reducing *presence* [1] and *immersion* [2]. Games like *Horizon: Zero Dawn*[1] overcome this problem by justifying the overabundance of head-up display (HUD) elements with an in-game excuse (in this particular case: a high-tech tracking device the main character wears). However, this is not always possible for every video game.

Through this article, we describe our study on how to guide a player in a virtual environment exclusively using audio. Our premise is that we can reduce the need of a cluttered GUI while retaining immersion and player performance, by letting participants firstly choose their preferred sound attributes and then adapting the soundtrack to these preferences.

In section 2, we start by analysing previously published work on adaptive music and player navigation using sound clues. In the next section, the experiment used to validate our proposal is described; its results are later summarized in section 4. Lastly, in sections 5 and 6, we include a brief discussion on the implications of our findings and several conclusions about how these ideas could be used in the design of a commercial adaptive music system for video games.

[1] https://www.guerrilla-games.com/play/horizon

2. PLAYER NAVIGATION AND ADAPTIVE MUSIC

The idea behind this article emerged from our previous work [3], which suggested there could exist a correlation between variations in the basic elements of a certain soundtrack and player decisions during an interactive experience: harmonic, high-pitched melodies seemed to attract users more efficiently than cacophonous, low-pitched ones. However, participant's reactions and behaviour varied greatly depending on the result each user achieved during the Bartle test [4]: certain groups of subjects were attracted to musical attributes which did not work as a lure for others. This led to the conclusion that personal auditive preference is important when using sound to orient players in video games.

We were also inspired by previous research with blind people [5], which acknowledges the existence of a conceptual level, in addition to a perceptual one, in the learning process associated with scouring an unknown environment in search for clues that allow to build mental, 3D "maps". This kind of perspective is of utmost importance for our research, because we base our work on the existence of culturally attained categories which relate to formal auditive parameters.

2.1. An adaptive music system

Adaptive music used in video games usually consists of an atmospheric, non-spatialised soundtrack which changes in response to specific events taking place in the virtual world. Said changes can happen procedurally or be previously scripted.

Due to the existence of very different social groups in terms of musical perception, we decided to build an adaptive, live music system, so as to respond in real time to player decisions while they play video games. This system is called LitSens [6, 7], and works by automatically combining short fragments of music composed by a human. LitSens was used in the present research as an audio foundation for the game we utilised in the experiment, which is described in section 3.

Our intention was to parameterise a series of basic musical attributes, so as to be able to modify them in real time with ease and efficiency. Our approach was similar to those of systems like ANTESCOFO [8], which go beyond pitch in terms of simple audio descriptors.

Furthermore, LitSens approaches adaptive music from an emotional perspective. The idea behind it is to adapt to player's emotional responses, in a way that allows a game designer to provide an adaptive soundtrack without taking into account every possible interaction or outcome. Wallis, Ingalls & Campana [9] approach this problem in a very similar way: they extract certain components, such as valence and arousal, from common emotional models, and apply them to music generation in real time.

2.2. Guiding players in virtual environments

The problem of guiding player movement in a virtual environment is a common issue in game design. Some authors, like Milam & El Nasr [10], have established a taxonomy of design patterns in 3D games, aiming to standardise these strategies, but academic work on player orientation through sound is scarce. Additionally, sound is not usually even taken into account when building these guiding techniques: none of the five patterns proposed by Milam & El Nasr make explicit use of sound. This opens an unexplored field of possibilities for game designers, who usually rely on visual clues.

It is not uncommon, however, to find alternative guiding techniques based on a video game's narratives. Earlier approaches, like the one presented by T. A. Galyean [11], rely on a path established by a narrative, which the user must follow to keep up. Recent commercial video games, like *The Stanley Parable*[2], *Dear Esther*[3] or *Gone Home*[4] all rely on narrative elements (e.g.: the voice of a narrator) to guide players to the next goal or important milestone. However, these techniques also rely on small or highly controlled virtual environments. Entangled paths or huge, open worlds require a different approach, and sound could be the key to solve the problem of subtle navigation assistance.

2.3. Auditive preference and meaningful variations

Additionally, Eisenberg & Forde[12] show that it is possible to establish a series of simple predictors, like creativity, complexity or technical goodness, which explain the variations in preference during a human evaluation of music. Though people's musical taste or preference is commonly measured and evaluated with musical genres in mind [13], we are interested in modifying simple auditive features, which allow for a more flexible approach and are consistent with an adaptive music system such as LitSens. We used complexity, pitch and rhythm as the three modifiable attributes during our experiment, as will be explained in section 3. This decision is consistent with previous uses of musical complexity (in and out-of-key notes, harmonic versus dissonant layering) [14], pitch (high and low tone) [15] and rhythm (slow or fast) [16] to produce perceptible changes when listening to audio fragments. The technique we used to increase complexity was simply to introduce layers of sound –formed by out of tune intervals and dissonant chords– that disrupted the harmony of a base track. This can be appreciated when comparing the two spectrograms depicted in figure 1. Pitch and rhythm modifications were made without adding any layer to the base mix; instead, we simply modified those values in real time for the whole track using commands from the game engine.

As for what makes a sound "stand out" over others, a very common opinion, based on classic works by Fletcher & Munson (the famed Fletcher-Munson curves) [17] is that a higher pitch – around 2000 and 5000 Hertz (Hz)– will usually dominate a mix in terms of perceived loudness. However, it has been known for a long time that listener's perception of several auditive attributes, included tone dominance, can be influenced by many different factors. Regarding pitch perception, in certain conditions[18], lower frequencies can be dominant. In the context of this research, dominance is a determinant factor when identifying and following spatialised sounds.

3. EXPERIMENT DESIGN

The following experiment had the objective of exploring the relationship between the presence or absence of adaptive music in a video game and player performance while solving a labyrinth-like orientation puzzle. It also measured the level of coherence between users' perception of sounds and their actual response to them.

3.1. Design

Before starting with the experiment, all participants were randomly distributed in two groups: A and B. Initially, both groups had the same size (N = 17), though group A lost a subject due to hearing health problems. Throughout the experiment, only two persons were in the area at a time: one participant and one test supervisor. There were four differentiated phases in every session: SAM test, attribute selection, game playing and sociological survey.

Subjects from group A started by taking a Self-Assessment Manikin (SAM) test [19, 20] about three pairs of sounds. Each pair was played consecutively, and had a strong relationship with one of the basic categories used to classify sounds in our test-bed game. The sounds in every pair represented the two opposed concepts for each of the following categories, presented in order in the test: tone (low-high), structure (simple-complex) and rhythm (slow-fast). The differences between the sounds of each category were big enough to be easily noticeable, as can be seen in figure 1, and during the test all sounds were evaluated separately after listening to each pair, in order to compare them.

The SAM test was passed in its 9-point scale version, by means of a digital form which contained all three measurements: emotional valence, arousal and dominance. This test uses the Semantic Differential [21] as a basis, and simplifies it. Thus, *emotional valence* measures "pleasure", and is strongly related to bipolar adjective pairs such as unhappy-happy, annoyed-pleased, unsatisfied-satisfied, melancholic-contented, despairing-hopeful or bored-relaxed. *Arousal*, on the other hand, is related to pairs like relaxed-stimulated, calm-excited, sluggish-frenzied, dull-jittery, sleepy-wide awake, unaroused-aroused. Lastly, *dominance* is related to adjectives like controlled-controlling, influenced-influential, cared for-in control, awed-important, submissive-dominant and guided-autonomous.

Subjects from group B were given the same test, but only evaluated one sound. This sound contained the default audio played by their version of the game, classified as: slow, low and simple. This evaluation was not taken into account later and it was performed to give the subject of this group the same insight than the subjects in group A about the auditive nature of the experiment, in order to avoid possible bias.

[2]https://www.stanleyparable.com/

[3]http://www.dear-esther.com/

[4]https://www.gonehome.game/

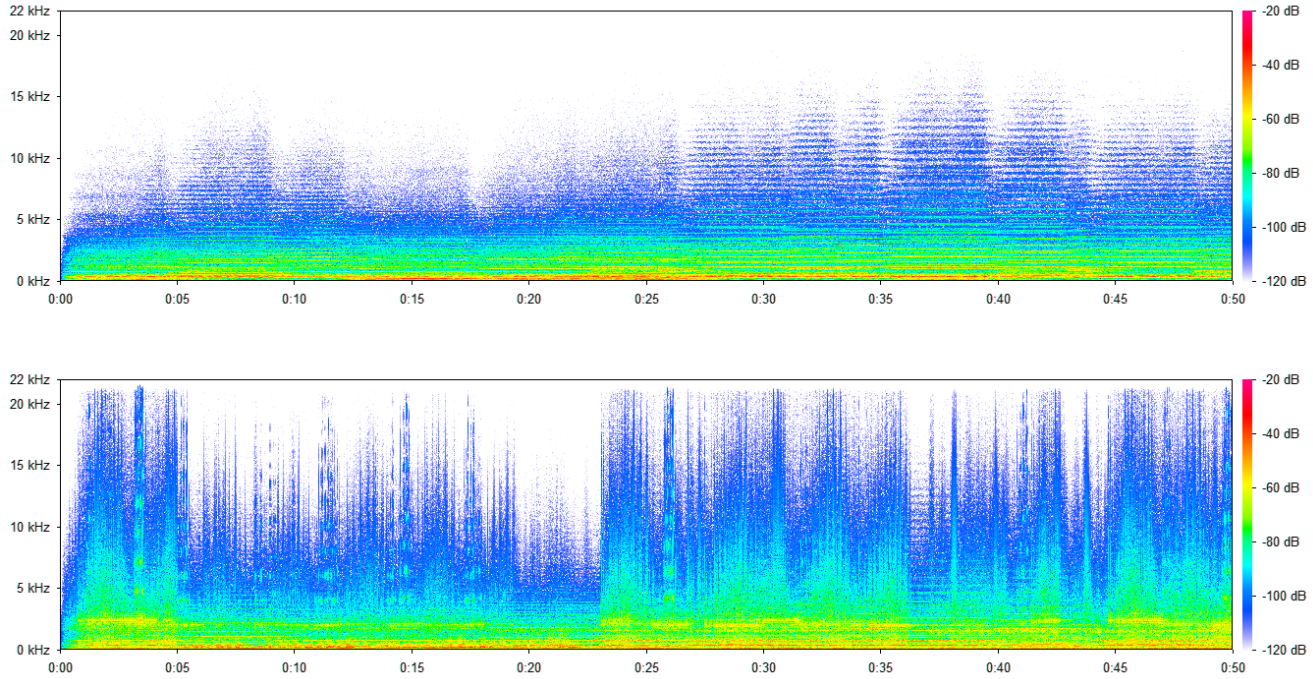

Figure 1: *Spectrogram of simple (top) and complex (bottom) variations of the same sound.*

Once finished with the test, subjects from both groups had to launch our software, which ran as a full-screen computer game developed with Unreal Engine 4 [5].

People in group A (experimental group) were asked to select, from an in-game menu with three categories (rhythm, tone and structure), the attribute for each of them (slow-fast, low-high simple-complex), which, from their point of view, would make a sound stand out over the rest. Thus, a total of 8 final outcomes were possible. People in group B (the control group) were not given this option, and played with default audio. No sound clues were included to help users from group A decide: the only previous reference was the SAM test. This was done in order to evaluate the coherence between subjects' perception of what sound suits them better and actual performance produced by their selection.

For group A, a personalized level was loaded after their preferences were specified. For group B, the level loaded with the default sounds (low tone, slow rhythm, simple structure). Said level consisted of a three dimensional labyrinth, played from a first-person perspective. From a logical standpoint, however, its structure can be considered two dimensional; it is depicted as a map in figure 2.

Players could move and look around using a keyboard (WASD keys) and a mouse. Every user was told to look for and recover a total of three statuettes inside this labyrinth, as quickly as possible. Elapsed time and number of statuettes recovered were shown on the screen permanently to keep the player informed at any time about his goal. The only way to recover a statuette was to step on it. Every time one of them was picked up, a measure of total elapsed time was stored in a log file. At the end of each session, this log was retrieved and tagged with the correspondent participant number. From now on, we will call the tree time measurements as follows, for convenience: t_1 (first statuette), t_2 (second statuette) and t_3 (third statuette, or total time).

Every statuette emitted a spatialised, monophonic music track which blended with a base stereophonic soundtrack. The base soundtrack was a low, synthetic drone, with no variations in tone or intensity. For users in group A, the emitted track was modified to adapt to their specified preferences in musical attributes. For users in group B, the track was always the default one. Once recovered, the statuette stopped emitting sound in every case, by means of a 2 second linear fade out. If the spatialised audio track was received by the camera listener through a wall, a low-pass filter with a cutoff frequency of 900 Hz was applied.

When all three objects were recovered, the game ended and the application was closed. After finishing with the game, every subject from both groups had to take a brief test to determine their sociological profile. Data retrieved included: age, sex, country of birth, level of education completed, presence of hearing problems, fondness for music and sound and performance when playing video games.

Subjects were also asked if sound was useful when trying to find the three statuettes inside the virtual labyrinth. Results from this question constitute a variable we named "help index" (h_i). A Likert 5-point scale [22, 23] was employed for this and all questions requiring gradation, except for the SAM test, where a 9-point scale was utilised.

Two leaflets with instructions were created, one for each group. Every subject had to read only the pertinent one while waiting to begin. These documents contained a detailed description of all actions every user would have to take during the experiment. Brief instructions on how to listen to the sounds and how to take the SAM test were included, as well as keyboard and mouse controls for the video game. All users were also told it was of utmost

[5] https://www.unrealengine.com/en-US/what-is-unreal-engine-4

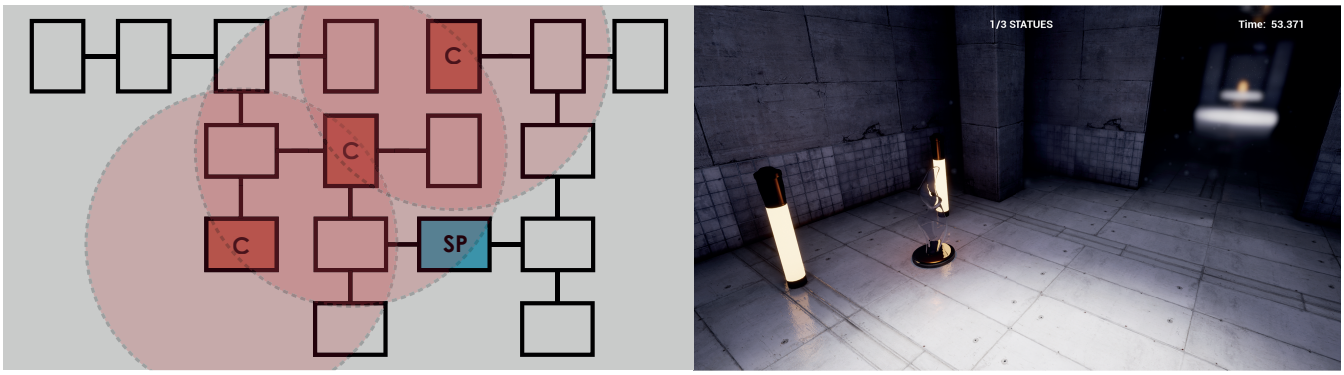

Figure 2: *Diagram showing the layout of the virtual environment utilised during the experiment. There is a starting point (SP) and three collectibles (C) in the form of statuettes. Red circles represent the area of influence of each sound. Walls applied occlusion through a low-pass filter, not depicted in the diagram. A capture of the game is also shown on the right.*

importance to complete the level in as few seconds as possible, and that they had to find three small statuettes to do so. The only difference between "A" and "B" versions was the lack of explanation on how to evaluate pairs of sounds (since this was not necessary for group B).

To evaluate results from both the 5 point Likert scales and the 9 point SAM scales a parametric, unpaired test (Student's t test) was utilised.

Finally, after finishing with the experiment, all subjects from group A were asked to explain, in their own words, the reasons for their attributes selection.

3.2. Hypothesis

Our hypothesis was that a statistical difference may be found between the two groups of users (A and B), in terms of performance (measured in total time, t_3), with the conditions established above. Our independent variable is the presence or absence of a preference selector at the beginning of the experiment that influences music played in the game. We also aimed to find a relationship between the initial selection of auditive features (available to participants in group A only) and t_3.

3.3. Demography

There existed two prerequisites participants had to meet so as to take the experiment: the ability to hear properly and having played at least one video game of the first person shooter (FPS) genre. All subjects met these requirements, and were students (graduate and postgraduate) or worked as lecturers in the field of Computer Science.

In total, 33 subjects participated in our experiment, of which 16 were assigned to group A (composed of 14 males and 2 females) and 17 (with 15 males and 2 females) to group B. From the total number, 29 were born in Spain, and the other 4 were from Colombia, Bolivia, Switzerland and Venezuela. All of them were native speakers of Spanish, which was the language used throughout the whole experiment. Also, they shared similar cultural features, and all but one had lived most of their lives in Spain.

Average ages in groups A and B were similar: 23,438 (A) and 24,059 (B) years. The mode was 18 in both cases, as most participants were first-year students.

68.8 % of the participants were undergraduate students, whereas 6.3 % were studying a master's degree at the moment. The rest were Ph. D. students (12.4 %) or university professors and researchers (12.5 %).

When asked if they played games frequently, most subjects in groups A and B answered positively, with a mode of 5 out of 5 in both cases, and a mean of 4.5 (A) and 4.412 (B). They also considered themselves good video game players, achieving a mode of 4 out of 5 for both groups and an average of 3.688 (A) and 3.824 (B). These scores were slightly lower when asking them if they were good with FPS games: the mode was 3 out of 5 in A and B, while the averages were 3.5 (A) and 3.353 (B).

As for self-evaluation of their hearing proficiency, when asked if they have good hearing, the modes were 5 (A) and 4 (B) out of 5, and the averages, 4.125 (A) and 4 (B). Besides, when told to answer if they have a "good ear" for music, the mode was 4 out of 5 in both cases, and the averages were 3.688 (A) and 3.353 (B).

Musicians were selected and distributed evenly between groups, with a total of 2 in each one. This was done to avoid possible bias due to their knowledge of music and audio, and they were the only participants which were not randomly distributed. This process occured before starting with the experiment, and the affected participants were unaware of it.

This leaves us with a surveyed sample which has a very good perception of their own hearing, but with an average-to-neutral confidence in their "musical ear".

4. RESULTS

The results of the previously described experiment (and its associated survey) point to several statistically significant differences between groups A and B in terms of both performance and self-assessment.

Subjects from group A achieved a total average time of completion (t_3) of 78.108 seconds, whereas participants in group B took an average of 132.987 seconds. The median in group A is 75.694, while in group B is 100.668. The lack of similarity between average and median times in group B can be explained by the presence of two clear outliers (as seen in figure 3), who completed the level in 369.250 and 367.020 seconds respectively. A parametric analysis of these results can be consulted in Table 1.

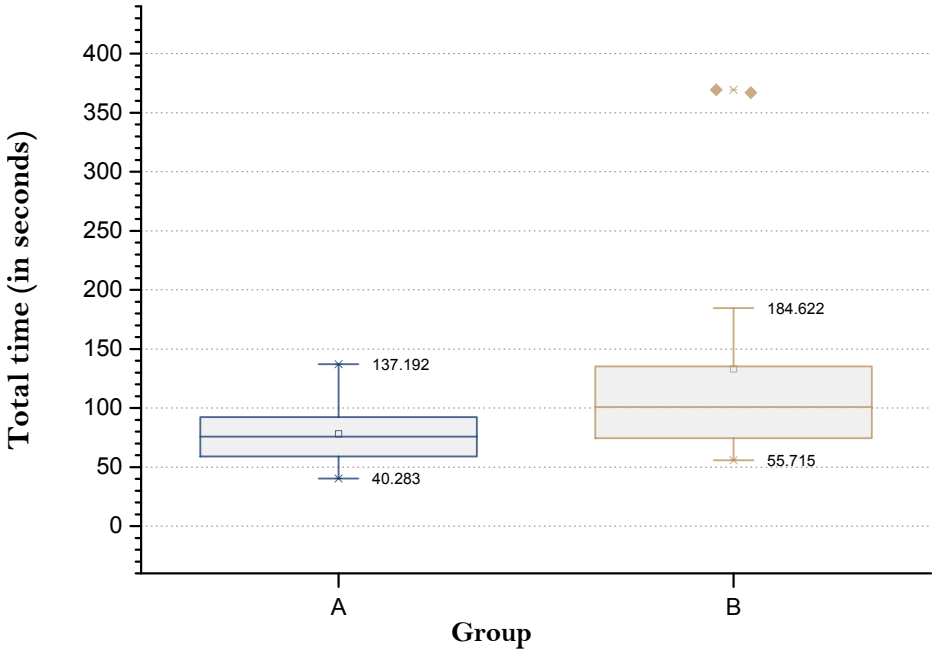

Figure 3: *Difference between groups A and B in total time (t_3).*

Table 1: *Student's t-Test for total time (t_3) in groups A and B.*

Group	A	B
N	16	17
Mean	78.108	132.987
Standard deviation	27.908	96.090
Two-tailed P value:	0.0356	

Table 2: *Student's t-Test for h_i values in groups A and B.*

Group	A	B
N	16	17
Mean	3.56	2.47
Mode	5	1
Standard deviation	1.46	1.59
Two-tailed P value:	0.0484	

Because time was measured when every statuette was picked up, not only total elapsed time gave an important insight about player behaviour during the experiment. It is also quite illustrative to look at how the difference in average time between the two groups increases as every object is taken. t_1 had an average value of 22.068 for group A, and of 25.637 for group B, so the difference between means equals 3.569 seconds. t_2 has an average value of 41.854 for group A and of 53.398 for group B, producing a difference of 11.544 seconds. Lastly, t_3 presents the biggest difference: 54.879 seconds.

As can be seen in Table 2, when it comes to the help index (h_i), there are also statistically significant differences between groups. Out of 5, group A has a mean of 3.56, while group B scores 2.47. The mode is particularly enlightening in this case: 5 in group A and 1 in group B.

There does not exist a strong statistical relationship between t_n and the initial selection of auditive features, which was only possible for members inside group A, as Table 3 shows. "High" (9) "fast" (9) and "simple" (11) were the most common options, however.

It is also worth mentioning that the attribute "complex" was the least selected (only 5 users chose it). However, this same attribute obtained the highest dominance score during the SAM test, with an average of 5.875 and a mode of 7 out of 9. It also had the highest excitement score, averaging 5.688 and with a mode of 7 out of 9. Additionally, general results from the SAM test were not consistent with player selections of attributes before playing the game (see Table 4).

It was not uncommon that, in spite of considering a complex sound more dominant, players chose a simple one instead before playing the game. Opinions around the very concepts of valence, dominance or arousal when it comes to detecting spatial sound were varied, and every user ended up choosing what appealed to them most. For example, when asked about the reason for their selection, 3 users mentioned "storms" or "thunder" as a reason for considering low tones useful when trying to orient themselves. They found those sounds "easy to track", "full of energy" or "very deep". The rest gave similar reasons for their decisions, based on personal experiences.

Table 3: *Features selected by group A participants and total time achieved (t_3).*

t_3	**Tone**	**Rhythm**	**Complexity**
40,283	High	Fast	Simple
42,159	High	Fast	Simple
45,318	High	Fast	Simple
57,027	Low	Slow	Simple
60,738	Low	Fast	Simple
66,320	Low	Slow	Simple
73,127	Low	Slow	Complex
73,483	High	Slow	Simple
77,904	High	Fast	Simple
81,639	High	Fast	Simple
84,315	High	Fast	Complex
92,305	Low	Slow	Complex
92,315	Low	Slow	Simple
95,468	Low	Fast	Complex
130,129	High	Fast	Complex
137,192	High	Slow	Simple

Table 4: *SAM test results in 9 point scale for variations of the same sound.*

Attribute	**SAM scale**	**Average**	**Mode**
1. High tone	Valence	5.938	7
	Arousal	3.625	2
	Dominance	4.5	5
2. Low tone	Valence	4.697	3
	Arousal	3.152	2
	Dominance	4.727	3
3. Simple structure	Valence	5.688	4
	Arousal	3.563	3
	Dominance	4.188	3
4. Complex structure	Valence	3.375	5
	Arousal	5.688	7
	Dominance	5.875	7
5. Fast tempo	Valence	6.063	7
	Arousal	5.25	7
	Dominance	5.188	5
6. Slow tempo	Valence	5.375	6
	Arousal	3.438	3
	Dominance	5.063	5

5. DISCUSSION

Considering the information retrieved through the previously explained results, we can extract a series of final deductions. First, the independent variable (the presence or absence of an attribute selector affecting music in the game) seems to be statistically related to the difference in total time (t_3) obtained by users during the experiment.

Besides, subjects in group A had a higher result in h_i, which means they perceived music as a helper more than participants in group B. Precedents for this effect have not been found in pertinent academic literature.

We have also observed that $t_n - t_{n-1}$ greatly increases with every measurement –whenever a statuette was gathered. This is inversely proportional to the number of statuettes present in the map. It is reasonable to think the amount of time elapsed in finding a statuette can increase when their remaining number is lower, because the probability of finding them by chance is also reduced. A need to backtrack and search more thoroughly also emerges when there are fewer objects to retrieve. However, the increasing variation in average t_n between groups A and B (see section 4) points to another, more important correlation. If we take into account that both prototypes (A and B) were identical except for the personalized music, it is possible to link the differences in mean time to the differences in audio.

Moreover, there were some counterintuitive aspects in the results. For example: the lack of consistency between SAM test results and player preference when selecting attributes inside the prototype could be happening due to multiple reasons. We have not retrieved enough information during our experiment to give a clear response to this particular matter, but several new and interesting lines of research are open as a result. Our main hypothesis for this unexpected behaviour is that the mere act of selecting sound attributes while already playing the game may not be in line with the mental state of the subject when answering the SAM test. While the test is a more relaxed experience, which is not limited by time constraints, the video game asks players to concentrate much more, and gives them a clear goal. As a consequence, it is possible that different attributes are found dominant in these different contexts, creating the mentioned variations in the results.

Another possible reason is that users learned to better identify dominant attributes through the duration of the SAM test, taking into account the specific variations in complexity, pitch and rhythm presented to them. This would mean the first answers would be less informed than the last ones, and that their decisions inside the final selector would imply a previous and meticulous "weighting up" of every possible option.

An appropriate new line of experimentation would involve distributing subjects in two groups in which the order of the test and the attribute selection would be inverted. Also, the SAM test only accounts for emotional scales (valence, arousal and dominance), and different measures could be needed to determine how easy to track a sound is for different persons, as sounds traditionally considered to be more dominant may not be easier to track for all users, and personal preference could be more important than dominance when it comes to finding sound sources in virtual environments.

Previous statements aside, user capacity to select attributes and user performance are, nevertheless, statistically related in our results. Consequently, we can state the mere ability to choose correlates with a lower average time of completion in group A, when compared to group B.

Other issues exist concerning data recovery and user distribution. For example: if we follow the Central Limit Theorem [24], our presumption of normal distribution would only solidly apply to groups with a number of participants (N) equal or greater than 30. However we want to note that our t_3 histogram forms a bell-like curve in both groups, even with less data, and the confidence interval of the mean is high enough (95 %) to trust the results. Nonetheless, a bigger sample would be needed to increase the reliability of the outcome. A similar problem is also the lack of women in our sample (only 4 out of 33 participants), which produces a genre bias and makes our retrieved data only strictly applicable to men. We aim to solve these predicaments in future iterations of this research.

6. CONCLUSIONS AND FUTURE WORK

The most relevant conclusion we can extract from the present research is the influence the mere act of selecting preferred attributes has over player performance when solving our 3D labyrinth. This effect, whilst somewhat predictable, was not verified in the past in any other research, so we open the path to further explore the consequences this causal relationship has in user behavior. For example, we observed that user preferences when selecting sounds differ from the ones chosen in the SAM test, so a future experiment would be necessary to analyse the rationale of this behavior.

Additionally, a similar experiment, based on player orientation, but without any kind of spatialisation or multichannel audio (that is: playing sound in mono in all channels), may help us elucidate if there are purely musical attributes that can make players take one path or another by themselves.

The increase in performance achieved when using adaptive, spatialised music may also make a preference selection system like the one we propose useful in different 3D environments where the inclusion of a GUI is not an option (such as virtual reality interactive experiences).

We noted also the surprising variety in attributes selected by subjects in group A. This attribute variety suggests the existence of a very complex population in terms of auditive preference when it comes to player orientation. Analysing this aspect in a bigger population and in more detail may prove useful for understanding what a "dominant" sound is in this context.

Another step we would like to take in the future to further validate our system would be to include LitSens in a commercial first-person video game and test whether we can guide players in bigger, more complex virtual environments.

Also, the development of an intelligent system integrated in LitSens, as a way to adapt to player musical preferences without a previous test, might improve immersion while reducing even more the amount of GUI elements needed. As the mentioned system already has the capacity to produce continuous adaptive music, only a new logic for the automatic selection process should be needed.

Lastly, it would also be useful to research how the level of presence achieved by systems which rely on GUI elements to guide a player varies when compared to systems using only sound to achieve similar results.

Consequently, the next experimental iteration for LitSens would have to take place in two separate steps: On one hand, the development of an intelligent system which would take into account player actions and camera movement to evaluate how users' context affects auditive predilections and consequently how this preferences impact performance.

On the other, an experimental validation with three groups of users ($N \geq 30$), which ought to include an implementation of the system in a commercial video game with open world environments and a presence test for all subjects (such as the Temple Presence Inventory [25]). Again, participants from group A would have access to adaptive, spatialised music, while group B would listen to a default, non-adaptive but spatialised soundtrack. Group C would listen to adaptive audio without any spatialisation (mono). Performance would be tested in terms of time when completing a navigation-related task, and camera movement would also be recorded.

We think that there is still much room for improvement in the field of intelligent management of sound systems for user navigation, especially when compared to the current state of image-based systems, which are much more developed. Audio is a less explored field in terms of semantic guidance, but it could substantially improve immersion and presence in virtual environments and be a useful tool for game designers. This is especially true when developing first-person or virtual reality experiences which cannot rely as heavily on GUI.

7. ACKNOWLEDGMENTS

This work has been partially supported by project *ComunicArte: Comunicación Efectiva a través de la Realidad Virtual y las Tecnologías Educativas*, funded by *Ayudas Fundación BBVA a Equipos de Investigación Científica 2017*, and project *NarraKit VR: Interfaces de Comunicación Narrativa para Aplicaciones de Realidad Virtual (PR41/17-21016)*, funded by *Ayudas para la Financiación de Proyectos de Investigación Santander-UCM 2017*.

We would also like to acknowledge the funding provided by Banco Santander, in cooperation with Fundación UCM, in the form of a predoctoral scholarship (CT2716 - CT2816) which contributed to the development of this research.

8. REFERENCES

[1] W. Barfield and S. Weghorst, "The sense of presence within virtual environments: A conceptual framework," *Advances in Human Factors Ergonomics*, vol. 19, pp. 699, 1993.

[2] C. Jennett, A. L. Cox, and P. Cairns, "Measuring and defining the experience of immersion in games," *International journal of human-computer studies*, vol. 66, no. 9, pp. 641–661, 2008.

[3] M. López Ibáñez, "Bartle Test Applications in Narrative Music Composition for Video Games," in *I Congreso Internacional de Arte, Diseño y Desarrollo de Videojuegos*, Madrid, 2015, pp. 1–13, ESNE.

[4] R. Bartle, "Hearts, Clubs, Diamonds, Spades: Players who suit MUDs," *Journal of MUD research*, vol. 6, no. 1, pp. 39, 1996.

[5] O. Lahav, "Improving orientation and mobility skills through virtual environments for people who are blind : Past research and future potential," *International Journal of Child Health and Human Development*, vol. 7, no. 4, pp. 349–355, 2014.

[6] M. López Ibáñez, N. Álvarez, and F. Peinado, "Towards an Emotion-Driven Adaptive System for Video Game Music," in *ACE 2017*, London, 2017.

[7] M. López Ibáñez, N. Álvarez, and F. Peinado, "LitSens: An Improved Architecture for Adaptive Music Using Text Input and Sentiment Analysis," in *C3GI 2017*, Madrid, 2017.

[8] A. Cont, "ANTESCOFO: Anticipatory Synchronization and Control of Interactive Parameters in Computer Music.," *Proceedings of the 2008 International Computer Music Conference, Belfast, Northern Ireland*, pp. 33–40, 2008.

[9] I. Wallis, T. Ingalls, and E. Campana, "Computer-Generating Emotional Music: the Design of an Affective Music Algorithm," *Proceedings of the 11th International Conference on Digital Audio Effects*, pp. 1–6, 2008.

[10] D. Milam and M. S. El Nasr, "Design Patterns to Guide Player Movement in 3D Games," *Proceedings of the 5th ACM SIGGRAPH Symposium on Video Games - Sandbox '10*, vol. 1, no. 212, pp. 37–42, 2010.

[11] T. A. Galyean, "Guided navigation of virtual environments," in *Proceedings of the 1995 symposium on Interactive 3D graphics - SI3D '95*, New York, New York, USA, 1995, pp. 103–ff., ACM Press.

[12] J. Eisenberg and W. F. Thompson, "A Matter of Taste: Evaluating Improvised Music," *Creativity Research Journal*, vol. 15, no. 2, pp. 287–296, jul 2003.

[13] M. Grimaldi and P. Cunningham, "Experimenting with music taste prediction by user profiling," *Proceedings of the 6th ACM SIGMM international workshop on Multimedia information retrieval - MIR '04*, p. 173, 2004.

[14] E. Fedorenko, A. Patel, D. Casasanto, J. Winawer, and E. Gibson, "Structural integration in language and music: Evidence for a shared system," *Memory and Cognition*, vol. 37, no. 1, pp. 1–9, 2009.

[15] J. H. McDermott and A. J. Oxenham, "Music perception, pitch, and the auditory system," *Current Opinion in Neurobiology*, vol. 18, no. 4, pp. 452–463, 2008.

[16] R. F. Day, C. H. Lin, W. H. Huang, and S. H. Chuang, "Effects of music tempo and task difficulty on multi-attribute decision-making: An eye-tracking approach," *Computers in Human Behavior*, vol. 25, no. 1, pp. 130–143, 2009.

[17] H. Fletcher and W. A. Munson, "Loudness, Its Definition, Measurement and Calculation," *Bell System Technical Journal*, vol. 12, no. 4, pp. 377–430, oct 1933.

[18] R. J. Ritsma, "Frequencies Dominant in the Perception of the Pitch of Complex Sounds," *The Journal of the Acoustical Society of America*, vol. 42, no. 1, pp. 191–198, jul 1967.

[19] M. M. Bradley and P. J. Lang, "Measuring emotion: The self-assessment manikin and the semantic differential," *Journal of Behavior Therapy and Experimental Psychiatry*, vol. 25, no. 1, pp. 49–59, 1994.

[20] B. Geethanjali, K. Adalarasu, A. Hemapraba, S. P. Kumar, and R. Rajasekeran, "Emotion analysis using SAM (Self-Assessment Manikin) scale," *Biomedical Research*, vol. 2, pp. 1–1, 2017.

[21] A. Mehrabian and J. A. Russell, *An approach to environmental psychology*, The MIT Press, Cambridge, 1974.

[22] R. Likert, "A technique for the measurement of attitudes," *Archives of Psychology*, vol. 22, no. 140, pp. 55, 1932.

[23] G. Norman, "Likert scales, levels of measurement and the "laws" of statistics," *Advances in Health Sciences Education*, vol. 15, no. 5, pp. 625–632, 2010.

[24] M. Rosenblatt, "A Central Limit Theorem and a Strong Mixing Condition," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 42, no. 1, pp. 43–47, jan 1956.

[25] M. Lombard, T. B. Ditton, and L. Weinstein, "Measuring Presence: The Temple Presence Inventory," in *Proceedings of the 12th Annual International Workshop on Presence*, 2009.