Design and evaluation of a gesture driven wavefield synthesis auditory game

Francesco Grani
Aalborg University Copenhagen
Copenhagen, Denmark
fg@create.aau.dk

Razvan Paisa
Aalborg University Copenhagen
Copenhagen, Denmark
rpaisa11@student.aau.dk

Jan Stian Banas
Aalborg University Copenhagen
Copenhagen, Denmark
jbanas14@student.aau.dk

Iakovos Vogiatzoglou
Aalborg University Copenhagen
Copenhagen, Denmark
ivogia00@student.aau.dk

Stefania Serafin
Aalborg University Copenhagen
Copenhagen, Denmark
sts@create.aau.dk

ABSTRACT

An auditory game has been developed as part of our research in Wavefield Synthesis. In order to design and implement this game, a number of technologies have been incorporated in the development process. By pairing motion capture with a WiiMote new dimension of movement input was achieved. And by delivering sound via wavefield synthesis spatial audio technique an immersive and natural auditory landscape has been created in which players can move around and interact with a set of sounds to pursue the game goal.

We present in this work an evaluation study where the game was assessed.

Author Keywords

NIME, Spatial Sound, Wavefield Synthesis, Audio Game, Realtime Synthesis, Sonic Interaction.

1. INTRODUCTION

The evolution of audio technology allowed for new listening setups to be experimented and evaluated. Long gone are the days of Thomas Edison's phonograph in 1877. Without doubt a milestone in the history of audio engineering, Edison's invention was able to both record and playback sound, however spatial fidelity was rather underwhelming, as the entire process was monophonic. Notably, not long after phonograph introduction, in 1881, a stereophonic playback device called the théâtrophone has been proposed by Clement Ader. The principle was simple - two microphones were placed across the opera stage and the signal collected by them was output to a pair of telephone receivers, placed in the opera house's foyer [23]. Later extensive research in this field slowly lead towards the commercial use of stereophony [25]. For some purposes it has been enhanced with an addition of a central speaker - mainly in cinemas, due to large dimensions of the screens. In consumer grade applications, stereophony has started to become widespread in the late 1950s with the invention of methods to engrave two channels onto a vinyl disc.

At the same time, spatialisation of sound sources is an expressive tool that music composers had put into use since centuries. Dozens are the compositions of the 16th century Italian composer Giovanni Luigi da Palestrina that make use of spatial distribution of musicians. With the rise of the era of electronic music during the second half of the 20th century, the number of composers who pushed the boundaries of the available techniques in order to pursue their creative needs in terms of spatial sound just increased, often leaving commercial solutions behind the "brute force" *ad-hoc* methods adopted by composers and their sound technicians (just to mention few cases: Karlheinz Stockhausen's *Gesang der J́unglinge* (1955), Varese's *Poeme Electronique* (1958)). In particular cases artistic needs ended up in the construction of dedicated venues such as the Acousmonium, designed in 1974 by Francois Bayle to host spatialised sound concerts [12]. In most cases however the bridge between science and art has been very short, leading to various experiments in the field of recording and mixing techniques (e.g., [9]) that quickly brought us to a series of available techniques for the recording and reproduction of almost any desided sound field. Results achieved in sound spatialization techniques for systems of loudpseakers span today from stereo panning to more extended multichannel configurations, such as ITU 5.1 Surround [22], VBAP [20] and [21], DBAP [15], ViMiC [19], first and higer orders of Ambisonics [14], [13], as well as this project focus, wavefield synthesis (WFS) [4], [5].

Of all the above mentioned techniques, however, only WFS (and its close relative ViMiC) let in principle the listener to perceive the same designed soundfield (and its virtual acoustic sources) in the same way and with the same auditory perspective from any point in space, since no "sweet spot" area affects the rendered soundfield. This peculiar characteristic makes WFS a privileged technique to be adopted in situations like the one described in this work, in which the listener is desired to have freedom of movement and to be completely able to walk around in a certain area, to interact with sonic events spatially distributed into a fairly large virtual acoustic space.

The primary concept explored in this project is the perception of distance, based on auditory cues only, in a synthetic acoustic space generated by wavefield synthesis. In a natural auditory environment, the listener is able to understand various auditory cues. In a virtual environment, most of the cues need to be mathematically reproduced in

NIME'16, July 11-15, 2016, Griffith University, Brisbane, Australia.

order to obtain a similar listening experience. In order to analyze how people perceive distance in a virtual auditory environment, a hardware and software environment was implemented to put under a unique flow of interaction user input and system feedback, coupling the wavefield synthesis system with a WiiMote and a visual motion tracking system.

2. BACKGROUND

The use of wavefield synthesis to recreate 3D sounds is not something new, having been pioneered in the late 80s at the TU Delft University, Netherlands [4]. In recent years, with the increase in availability of computational power the technology has gained a commercial interest, supported by hardware and software solutions such as the ones proposed by IOSONO [7] and Sonic Emotion [10],[18] as well as several open source engines available to control, simulate and render WFS [1], [2], [3], [24].

Before presenting the test scenario and results of this project, the following chapter will briefly present the basic principle of wavefield synthesis as well as a theoretical background on how the human hearing system perceives distance and spatial sound in general.

2.1 Basics of Wavefield Synthesis

The foundation of WFS lies on the theory concept of Christiaan Huygens: *Each point on a wave front can be regarded as the origin of a point source. The superposition of all the secondary sources form a waveform which is physically indistinguishable from the shape of the original wave front* [26].

The principle has been originally used to describe water and optical waves, and was first formulated for acoustics in 1988 at the TU Delft after being pioneering described in the 50s by Snow et al. [25]. A WFS system does require a large number of loudspeakers, placed as close as possible to the next one in order to create an array with as few discontinuity as possible. Each loudspeaker of the array corresponds this way to a secondary sound source and needs to be driven by a dedicated/independent signal thus requiring a large number of audio channels, equal to the number of loudspeakers; the signal for each channel is calculated by means of algorithms based on the Kirchhoff-Helmholtz integrals and Rayleigh's representation theorems [28], [29]. Due to the physical and software limitation, WFS systems are enduring several approximations, which introduce certain limitations and artifacts.

A first approximation needed to minimize complexity is to reduce the control of the sound field from a 3D to a 2D space (an horizontal -unlimited- plane). A second approximation consists into limiting the amount of secondary sources to a finite number (a finite set of loudspeakers); this approximation leads towards the consequence that the frequency range whereas a WFS system provides artifacts-free sounds gets reduced to the portion of the acoustic spectrum that is located below a threshold frequency, named "spatial aliasing frequency"; above this frequency artifacts will occur in the form of "ghost sound images". To cope with this limitation it is desirable to place the loudspeakers at the minimum possible distance -in our case 16.4cm, thus introducing a spatial aliasing threshold of 1048 Hz- and to design a sonic content which is not unbalanced towards hi frequencies. Another approximation consists on the fact that linear arrays of loudspeakers have a limited physical length and this generates what is called "truncation error", a phenomenon that limits the angles of incidence of sound sources in which a good result of WFS can be achieved. Further interferences can be introduced by the loudspeaker construction itself, as well as by the acoustics of the room in which the system is installed. An exhaustive description of WFS limits can be found in [27].

In wavefield synthesis technique it is usual to distinguish between three categories of sound sources that can be reproduced.

- Point sources: virtual sources that are placed anywhere outside the inner area of the loudspeakers array. These should be perceived as having the same virtual position when listened from anywhere in the inner area of the loudspeakers array.
- Plane waves: similar to point sources, they are placed behind the loudspeakers array but ideally at an infinite distance, thus their incident wavefront can be described as plane. These sources should be perceived identically at any position in front of the loudspeakers, as if they were a sound horizon (if the listeners walks in parallel to their wavefront, they should be perceived as if they were walking together with the listener).
- Focused sources: sound sources that are located in front of the loudspeakers array. Unlike sources behind the array, focused sources create a "grey" area between the virtual position in which they are placed, and the location of the loudspeakers. This can be described as pre-echo, and the sound can be perceived as not having a precise location; however the location of the sound source becomes immediately clear as soon as a listener "walks into it". A way to improve the resolution of focused sources is to adopt closed arrays of loudspeakers in the form of circular or rectangular arrays (as in this case).

2.2 Distance Perception

Human hearing has a very important role in our everyday orientation in space. Sounds usually convey information about their source and its location, and we have evolved to "decode" this information. The sound wave generated by a source is diffracted by its interactions with the head and external ears. The resulting changes in temporal and spectral characteristics of the sound provide us with cues about the localization of the source itself [17]. A number of different properties of physical stimulation have been thought to be potential cues to the perception of auditory distance. These cues may be generally divided between those that require to reach only one ear (monaural) to be informative, and those that require to be heard by both ears (binaural) [31]. The present description of these cues will therefore be classified on the basis of either monaural or binaural requirements.

Monaural cues are hints that contain information about the distance of an object to the listener. There are three popular monaural cues that will be described: sound intensity, spectral shape and the direct-to-reverberant ratio. If the first two do require a familiarity with the sound, the direct-to-reverberant ratio is, in theory, an absolute cue. In environments with sound reflecting surfaces, the ratio of energy reaching a listener directly (without contact with reflecting surfaces) to energy reaching the listener after reflecting surface contact (reverberant energy) decreases systematically with distance [32]. In general, as sound sources move away from an observer in a reverberant environment, the portion of sound energy directly reaching the observer's ears decreases, while the portion reaching the observer's ears after reflection from surrounding surfaces (with consequent delay) increases [16]. This is called the direct-to-reverberant ratio. In indoor environments, change in direct-to-reverberant energy ratio primarily follows the 1/dist ratio on the direct portion of the sound field, since the energy in the later arriving reflected portion of the sound field is relatively constant for varying source distance. The direct-to-reverberant ratio is a particularly interesting cue since it is, in theory, absolute; it does not depend on source intensity, nor to the familiarity of the listener with the source [8]. Direct-to-reverberant ratio is able to code a wide range of distances in many reverberant environments, and a listener can "learn" the acoustical environments thorough this cue, improving his/her own perception of sounds. This cue is also of interest because it has apparent limitations. It has been shown repeatedly [8] that the perceived distance of a sound source in a room is compressed; it increases virtually linearly with the distance from the source at short range, but converges to a certain limit when the source distance is increased further. This limit acts as a sort of "auditory horizon", which is, however, not constant but depends on the acoustic environment.

Binaural cues: when sound sources are located near to the listener the human auditory system uses both differences in intensity and time as spatial cues [6]. These differences are called Inter-aural Time Difference (ITD) and Inter-aural Level Difference (ILD): e.g. a sound coming from the left, will reach the left ear before reaching the right ear. ITD and ITL provide information for localization of sounds on the horizontal plane, while spectral spectral effects caused by head and torso diffractions provide information on the localization of sounds in the median (vertical) plane (an aspect this one that is not investigated by this study in which the control of wavefield happens in the horizontal plane).

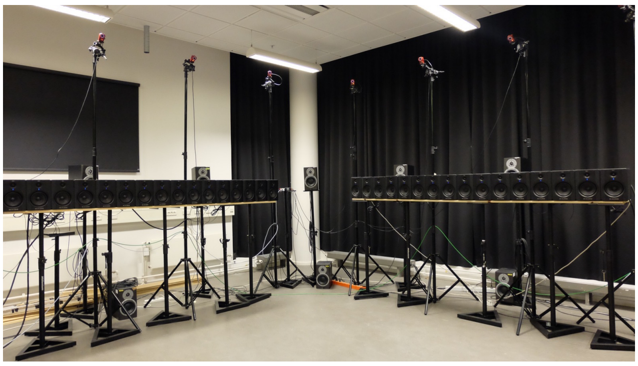

Figure 1: The playing area surrounded by the WFS arrays and OptiTrack Motion Capture system.

It is well known that our ability to perceive distances of sound sources depends on several different cues, and the auditory system likely combines information from multiple cues to produces stable distance percepts. Not all cues are equally effective in all circumstance. In the context of re-creating acoustic in virtual environments through wavefield synthesis, some cues can be seen as more influential than others. Since it is a controlled environment, an interesting cue to be investigated can be found in direct-to-reverberant ratio. Having control over the recordings and synthetic sounds, one can simulate different virtual sonic environments, which contain as many reflective virtual objects as desired. Besides direct-to-reverberation cue, another important one can be the spectral shape characteristic of the sound [30].

3. EXPERIMENT DESIGN

3.1 Overall Description

The aim of this project is to investigate how convolution reverb affects people's perception of distance in a wavefield synthesis setup environment. In order to achieve this, an auditory game prototype was developed and to keep the focus on auditory perception, players do play the game blindfolded. The style of the game is horror/survival and the user is exposed to several "enemies", which he/she needs to localize and eliminate by using a Nintendo WiiMote game controller to "throw" sounding objects towards them. There are three types of enemies with different mechanics and sonic characteristics that will be described in a next section. They all are created by using point sources and focused sources, and they are wither static or moving around or towards the player after they appeared in the virtual space around the player.

3.2 Impulse Response Reverb

The environment of the game resembles a commercial ship, thus a background ambience soundscape was designed containing sounds such as an air fan, water drops from a broken pipe, wind sound coming from outside the ship and rat squeaks. Acquiring the impulse response from a ship was essential, since this project relies on investigating the role of convolution reverb in distance perception for WFS. The Impulse Response was captured using the ESS (Exponential Sine Sweep) method [11], in a big metal ship owned by the Illutron Collaborative Interactive Art Studio (see `http://illutron.dk`). The following equipment was used in the process: a MacBook Air Laptop, a Dynaudio BM5 MK I speaker, a Rode NT2 omnidirectional microphone and a Focusrite Scarlett 8i6 audio interface. The recording and deconvolution was handled via the Apple Logic Pro X internal Impulse Response Utility. Seven different relative positions of the loudspeaker and microphone were tried during the process, finding the one with the microphone placed in one corner and the loudspeaker on the opposite wall of the room facing the microphone being the optimal.

3.3 Hardware and Software

Since the game was designed to allow the player to move freely in the area inside the array of loudspeakers, a shooting system has been implemented coupling a WiiMote with two motion capture markers captured by an array of 16 OptiTrack Flex 3 infra-red cameras. Of the two MoCap markers one had been placed on the player's shoulder and another one on the WiiMote. This way the direction of shooting could be obtained, as well as the position of the subject in the virtual space, allowing the player to shoot at any angle.

The cameras array is connected via USB to a PC running Windows 7. Cameras were calibrated and set up using OptiTrack NaturalPoint software, and broadcasted via VRPN protocol to a Unity3D instance running on the same computer. In Unity, a simple project was running, displaying two objects, each connected to one of the trackers. Unity3D was running as a local debug software as well as an interpreter from VPRN to OSC, since NaturalPoint cannot send OSC data. The Unity5 game engine, the WiiMote OSCulator 2.13.3 receiver and the WFSCollider sound engine software were all running on a Mac Pro computer (dual Intel Xenon 12 core processor, 64 GB DDR3 RAM).

The WFS audio stream is delivered from an RME MADIface USB interface to two DirectOut ANDIAMO 2 MADI AD/DA converters, each connected to 32 M-Audio BX5 D2 loudpspeakers (previously calibrated). In total the WFS system delivers sound trough 64 loudspeakers aligned one to the other (the distance between centres of two consequent speakers is 16.4 cm, which is barely the radius of the loudspeaker cone plus the enclosure thickness). The WFS system configuration consists of 4 arrays of 16 loudspeakers each, displaced to form a square of 4 by 4 meters inside which users can freely move. In Figure 1 can be seen half of the setup (OptiTrack cameras are also visible above the loudspeakers), the other half mirrors the visible one.

The wavefield synthesis had to happen in realtime after the user input and according to the enemies positions, to maintain the desired playability. The choice of WFS engine fell on WFSCollider, the audio spatialization engine for Super Collider developed by Wouter Snoei at The Game of Life Foundation (see `http://gameoflife.nl`). Beside the capability of rendering wavefield sound, WFSCollider also serves as an intuitive digital audio workstation (DAW) offering functionalities such as multi-track mixing, effect chains, auxiliary buses, featuring also an easy OSC control on every parameter, thus making it very suitable for the desired setup of this work. In WFSCollider sound sources are triggered and controlled in position and properties by control messages coming from Unity5 and OSCulator.

3.4 Sound Design

Three types of enemies were designed for the game, which will be described as Enemy 1, Enemy 2 and Enemy 3. The numbering represents the order of apparition. All these enemies spawn randomly at different locations, from three different "rings" or levels of distance. Enemy number one will always appear from the further area, while enemy number two will be appearing from the closest one, leaving enemy number three to appear from the mid one. See Figure 2 where E1, E2, E3 represents the three enemies; the rings represent the three different areas of distance where enemies are coming from; the square represents the physical space enclosed by the WFS array, and P represents a player.

The lifetime of each enemy is 1 minute. This limit is implemented to compensate for an issue encountered in the pilot tests: sometimes, the user cannot hit an enemy.

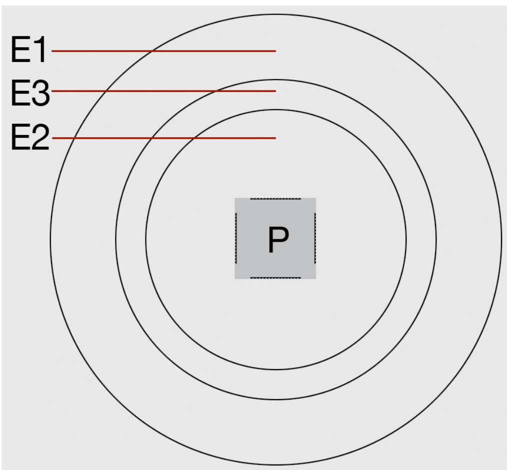

Figure 2: Map of the virtual space.

1. Enemy 1 slowly moves on a linear trajectory towards the player and tries to "hit" him/her, emitting a continuos flow of sound while it moves. E1 depicts a human dragging a heavy metal object and the most evident sound characteristic of this enemy is its slow footstep movement. Several sounds are used to create it: a pair of foot sounds alternating, a recorded heavy breathing sound, as well as the sound of a metal object dragged on a metal surface. All sounds are grouped together; as long as their virtual position is located outside the ring of speakers they are rendered as point sources, and when they get close to the player and "enter" the loudspeaker area, they become focused sources.

2. Enemy 2 position is static. E2 represents a woman who is breathing fast and sobbing while spinning a chain, its sound characteristics are then female screams and a swinging flail weapon sound. This enemy is immobile and it alternates short silences and sounds. Three sounds were used to create her, a chain links clinker, a recorded sobbing/ breathing sound and a vocal sound. Just as Enemy 1, these sounds are grouped together and if the enemy appears in the area behind the loduspeakers they are rendered as point sources, otherwise they are rendered as focused sources if E2 appears in the area inside the loudspeakers array.

3. Enemy 3 combines together some of the mechanics and sound characteristics of E1 and E2. Its position is static but every 20 seconds it spawns a series of moving distractive sounds, which travel around the virtual space where the player is, making it harder of the player to locate and eliminate him/her. E3 symbolizes a ward drum player, with a twist, and only one sound source is used to create it: a rhythmical uninterrupted drum loop. For the distraction sound, several male exhale sounds were used, being processed to sound like a wrath. These two sounds (E3 and its "distractors") combine together one continuous sound cue with a series of short sounds that appear and go. Just as the other two enemies, all the sounds are point sources when their virtual location is located outside the speaker area, and focused source otherwise.

3.5 Test Design

Each subject is introduced to the game mechanics by going through a training phase which consists of three stages, each lasting one and a half minutes and dedicated to set the player familiar with the relation between the gesture he/she has to perform (direction and force of the gesture) to "throw" a sound against an enemy, and the distance at which the sound is thrown. In this phase the subject is already blind-folded and is requested to locate and hit the virtual sound by "shooting" another sound with the WiiMote towards it, according to the subject's perception of how far the target sound is located.

The training sound to hit resembles a synthetic metronome beat and is located into one of the three circular areas visible in Figure 2; the player receives a sound feedback to understand if the shoot was good (the gesture was performed with the exact force needed to launch the sound into the desired area) or not. The sound to hit remains the same on all three stages but the distance increases from the inner to the outer circular areas as the stages progress. Once a participant has been familiarised with how the game works, the actual testing starts.

The real game/test comprises also of three stages, one for each of the three enemies. In each of the three stages the participant is exposed to eight instances of every enemy, four of these are presented with impulse response reverberation and four without, randomly assigned. The participants actions are tracked throughout the test and logged to files. The log entries include player position, collisions coordinates and timing, number of shots fired during each of the phases and number of enemies spawned and hit.

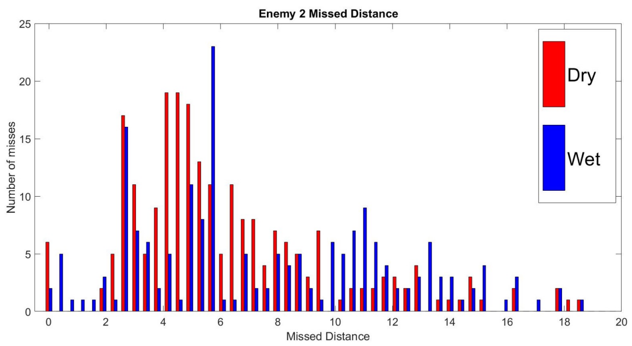

Figure 3: “Spatial precision”: missed projectile distance test results (wet/dry reverb conditions).

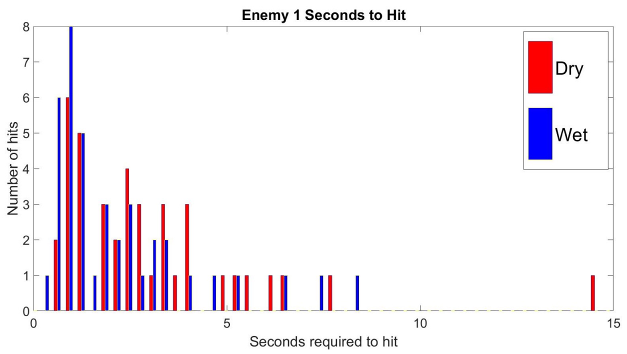

Figure 4: “Game performance”: time required to hit an enemy (wet/dry reverb conditions).

3.6 Experiment Results

The test was performed on 9 males and 1 female participants aged between 21 and 27, all of them reporting to have musical training and suffer no hearing loss. The results have been gathered in three main categories. These include a distance between the projectile impact position and a target enemy (namely “spatial precision”), the time required to correctly shoot an enemy and the total accuracy of shots (number of good shots versus bad shots). In each of these categories a paired t-test has been performed to verify whether or not the presence of reverberation (wet/dry parameter) had an influence on the participants’ achieved scores. This procedure was carried out three times, once for each of the enemies that the test participants were exposed to, thus getting a total of 9 tests. The paired t-test revealed that results provide no statistical significance required to determine whether or not entries related to the reverberant environment condition differ from those logged in the dry condition (= with no impulse response based reverberation - only the natural dry reverberation provided by the lab room in which the WFS system is placed). Among all the nine tests, only one yielded significant difference between two conditions - the one ran on missed projectile distance entries during an Enemy 2 phase (p=0,046). However, even in this case, the difference between mean values is equal to 20.1949-16.5106=3.6843, a relatively low number -Figure 3.

Statistical analysis of data does not bring a solid answer to the hypothesis that a difference is in place between the performance achieved in shooting at the correct distance in wet or dry reverberation conditions. Also the analysis performed on the number of seconds required to hit an enemy, shows no difference in all nine cases, so only one plot is here presented as an example of the results (Figure 4.), leaving further reflections to the discussion part.

4. DISCUSSION

The analysis of data shows no significant difference in the results performed with and without convolution reverberation, nevertheless it is worth mentioning that both the system used as a tool to perform the test, and the experiment design itself have possibly affected the outcome substantially. First of all, the gestural interface was commonly reported by subjects to be counter-intuitive and non-reliable and hence it can be partially blamed for an overall poor performance of the users (in terms of accuracy, time required to aim and average missing distance); moreover, this aspect raised frustration and distraction from the task. Consequently, participants tended to become tired towards the end of the test, which led to further deterioration of their score. The main reason behind this issue has been addressed as the delay between the motion capture system and the WiiMote input data flows. The stream of data from the MoCap computer, to the computer receiving the WiiMote data, is affected by a small lag, that causes incorrect reading on the users hand position in the moment when they trigger the WiiMote button to “throw” their sonic weapon. This small lag sometimes causes a wrong reading of the relative position of the two markers (the one placed on the player’s shoulder and on the WiiMote), which in the end can generate a wrong shooting angle. This error is more pronounced in users who perform a very fast and energetic movement with the WiiMote. This problem could be overcame by changing the shooting mechanism. Another solution to overcome the lag would be to redesign the data flow either making use of a single computer, or relying only on the WiiMote internal sensor data fusion to generate an accurate shooting direction.

The evaluation aspect of this project was revolving about the impact of convolution reverb in a WFS system, but this is not the only way to create artificial room simulations; different techniques could be adopted instead of a direct convolution of the sound sources: for future studies another option could be to model reverberation as four planar waves representing physical walls and fed with all the signals to be convolved. Also incorporating completely different approaches, such as Schroeder reverberators might be worth investigating. It is in the end worth mentioning that this project completely omits the proprioception aspect of the experience. Early tests suggest that the perception of the shooting hand might influence the shooting performance from player to player. A further experiment investigating this aspect could provide useful information in understanding the analysed data, as well as provide useful knowledge for designing interactions and interfaces for alike systems. At last, also sound design aspects could be affecting the results and be worth investigating more, since besides comparing moving sounds and static sounds, the sounds themselves embed different temporal and spectral contents which might affect subject’s perception.

5. CONCLUSIONS

This paper investigated the impact of convolution reverberation over the distance perception in a wavefield synthesis based auditory game: a system involving 4 arrays of 16 speakers, a motion capture system, and a WiiMote were used for this. An experiment was conducted that exposed 10 subjects to three types of enemy targets. There were 8 instances of each enemy, 4 were convolved with the impulse response of a ship room while 4 were not. A two sample t-test was performed on the data gathered. While the test results were mainly inconclusive, the miss distance for Enemy 2 was statistically proven to be influenced by the reverb status, accepting the null hypothesis, indicating that a dry sound sources were slightly easier for the participants to hit. Nevertheless, the platform used for this research is worth further development as it provides more possibilities for examining embodied interaction in a virtual spatial auditory environment. Also, besides the considerations on further possibilities of study on how to implement a more effective setup for the experiment purposes, another interesting way of exploiting it would be to include interaction between people, so that more users can interact with the environment.

6. REFERENCES

[1] J. Ahrens, M. Geier, and S. Spors. The soundscape renderer: A unified spatial audio reproduction framework for arbitrary rendering methods. In *Audio Engineering Society Convention 124*. Audio Engineering Society, 2008.

[2] M. Baalman. Application of wave field synthesis in the composition of electronic music. In *International Computer Music Conference, Singapore*, pages 1–4, 2003.

[3] M. A. Baalman. Updates of the wonder software interface for using wave field synthesis. *LAC2005 Proceedings*, page 69, 2005.

[4] A. J. Berkhout. A holographic approach to acoustic control. *Journal of the audio engineering society*, 36(12):977–995, 1988.

[5] A. J. Berkhout, D. de Vries, and P. Vogel. Acoustic control by wave field synthesis. *The Journal of the Acoustical Society of America*, 93(5):2764–2778, 1993.

[6] J. Blauert. *Spatial hearing: the psychophysics of human sound localization*. MIT press, 1997.

[7] S. Brix, F. Melchior, T. Roder, S. Wabnik, and C. Riegel. Authoring systems for wave field synthesis content production. In *Audio Engineering Society Convention 115*. Audio Engineering Society, 2003.

[8] A. W. Bronkhorst. Modeling auditory distance perception in rooms. In *Proc. EAA Forum Acusticum Sevilla*, 2002.

[9] J. M. Chowning. The simulation of moving sound sources. *Journal of the Audio Engineering Society*, 19(1):2–6, 1971.

[10] E. Corteel and T. Caulkins. Sound scene creation and manipulation using wave field synthesis. *Rapport technique, IRCAM*, 2004.

[11] A. Farina. Advancements in impulse response measurements by sine sweeps. In *Audio Engineering Society Convention 122*. Audio Engineering Society, 2007.

[12] B. François. Pour une musique invisible: un acousmonium. *Festival International du Son Haute Fidélité Stéréophonique*, pages 125–134, 1975.

[13] M. Gerzon. Surround-sound psychoacoustics: Criteria for the design of matrix and descrete surround-sound systems. *Wireless World. Reprinted in An anthology of articles on spatial sound techniques, part 2: Multichannel audio technologies. Edited by F. Rumsey. Audio engineering society, Inc, 2006.*, 1974.

[14] M. A. Gerzon. Ambisonics in multichannel broadcasting and video. In *Audio Engineering Society Convention 74*. Audio Engineering Society, 1983.

[15] T. Lossius, P. Baltazar, and T. de la Hogue. *DBAP–distance-based amplitude panning*. Ann Arbor, MI: MPublishing, University of Michigan Library, 2009.

[16] D. H. Mershon and L. E. King. Intensity and reverberation as factors in the auditory perception of egocentric distance. *Perception & Psychophysics*, 18(6):409–415, 1975.

[17] J. C. Middlebrooks and D. M. Green. Sound localization by human listeners. *Annual review of psychology*, 42(1):135–159, 1991.

[18] R. Pellegrini and C. Kuhn. Wave field synthesis: Mixing and mastering tools for digital audio workstations. In *Audio Engineering Society Convention 116*. Audio Engineering Society, 2004.

[19] N. Peters, T. Matthews, J. Braasch, and S. McAdams. Spatial sound rendering in max/msp with vimic. In *Proceedings of the 2008 International Computer Music Conference*, 2008.

[20] V. Pulkki. Virtual sound source positioning using vector base amplitude panning. *Journal of the Audio Engineering Society*, 45(6):456–466, 1997.

[21] V. Pulkki. Generic panning tools for max/msp. In *Proceedings of International Computer Music Conference*, pages 304–307, 2000.

[22] I. Recommendation. 775-1, multichannel stereophonic sound system with and without accompanying picture. *International Telecommunication Union, Geneva, Switzerland*, 1994.

[23] F. Rumsey. *Spatial audio*. CRC Press, 2012.

[24] W. Snoei. Wfscollider 2.2.1.

[25] W. Snow. Basic principles of stereophonic sound. *Audio, IRE Transactions on*, (2):42–53, 1955.

[26] T. Sporer. Wave field synthesis-generation and reproduction of natural sound environments. In *7th International Conference on Digital Audio Effects (DAFx-04), Naples, Italy*, 2004.

[27] S. Spors and R. Rabenstein. Spatial aliasing artifacts produced by linear and circular loudspeaker arrays used for wave field synthesis. In *120th AES Convention*. Citeseer, 2006.

[28] S. Spors, R. Rabenstein, and J. Ahrens. The theory of wave field synthesis revisited. In *124th AES Convention*, pages 17–20, 2008.

[29] E. N. G. Verheijen. *Sound reproduction by wave field synthesis*. PhD thesis, TU Delft, Delft University of Technology, 1998.

[30] F. L. Wightman and D. J. Kistler. The dominant role of low-frequency interaural time differences in sound localization. *The Journal of the Acoustical Society of America*, 91(3):1648–1661, 1992.

[31] P. Zahorik, D. S. Brungart, and A. W. Bronkhorst. Auditory distance perception in humans: A summary of past and present research. *Acta Acustica united with Acustica*, 91(3):409–420, 2005.

[32] P. Zahorik and F. L. Wightman. Loudness constancy with varying sound source distance. *Nature neuroscience*, 4(1):78–83, 2001.