Sound Hunter
Developing a Navigational HRTF-Based Audio Game for People with Visual Impairments

Sebastian Wolfgang Brieger
School of Computer Science and Communication
Royal Institute of Technology (KTH)
Stockholm, Sweden
brieger@kth.se

Abstract— In this article, a framework is proposed for designing 3D-based audio-only games in which all navigation is based on perceiving the 3D-audio, as opposed to relying on other navigational aids or imagining the audio as being spatial, where additional sounds may be added later on in the development process. To test the framework, a game named Sound Hunter was developed in an iterative process together with both sighted and visually impaired participants. The results indicate that the suggested framework might be a successful guidance tool when wanting to develop faster perception-based 3D-audio games, and the learning curve for the navigation was approximately 15 minutes, after which the participants navigated with high precision. Furthermore, with only small alterations to game menus and the iPhone's accelerometer function, both older and younger visually impaired people can navigate through 3D-audio environments by using simple hand movements. Finally, the results indicate that Sound Hunter may be used to train people's spatial hearing in an entertaining way with full experimental control. Two main factors seem to affect the learning curve for adapting to a foreign HRTF during virtual interactive gaming experiences; the adaptation to the navigational controls, and the experience of front/back confusion, where control adaptation is promoted by having a strong default setting with customizable sensitivity, and the experience of front/back confusion can be greatly reduced by introducing complex distance-dependent meta-level communication in synthesized sounds.

Keywords— Audio games, HRTF-synthesis, 3D-audio, game development, visually impaired, blind, HCI.

I. INTRODUCTION

Between 1996 and 2006, around 400 audio games had been developed, which is a very small number compared to visual computer games. The development teams were also small, usually consisting of one to four persons [1]. However the development has been more substantial over previous years, and the role of researchers, game developers, as well as sound designers has become more important in order to find more pleasant audio rendering techniques, as well as new exciting methods and technologies to use in audio games [14, 1]. There has been a rapid development of audio chips and 3D sound engines for computer games, and today, users as well as developers pay more attention to the audio content in computer games [7]. Sound is an expressive narrative medium, and sonic landscapes, or "soundscapes", may vary well be as immersive and engaging as powerful 3D-based graphical environments [7].

Still, even though audio games are developed more sophisticated these days, there is nonetheless very restricted access to an important part of the youth culture for people being visually impaired, and it could be argued that including this user group is of great importance, as it will aid their participation in society [1]. Accessibility to software applications is another area in which the visually impaired have more difficulties than sighted users. However, recent developments of various frameworks intended to make software applications more accessible has greatly aided the visually impaired, examples of which are Microsoft Active Accessibility, VoiceOver for the iPhone, as well as similar frameworks for Mac and Linux desktop environments [1]. By focusing more attention to game audio, new possibilities of designing games for people with visual impairments have emerged. However, this also requires that the games be developed with regard to their abilities and needs [7]. Games may also be used as a means to train the various senses and abilities in a person. For example, handicapped children may benefit largely both from using music [8] and computer games in order to aid their psychomotor and cognitive development [15, 1].

Despite the growing focus on audio in computer games, the audio content of mainstream computer games is still largely underdeveloped in comparison to the visual content [7]. This also applies to games being completely based on audio, which are extremely rare compared to audio-visual games. Almost all audio-only games on the market today are made for PC computers, where some of the popular titles are SuperDeekout [5] and Terraformers [18]. Still, as the number of blind iPhone users today is over 100 thousand and rapidly growing [23], it therefore seems as if there is a significant and highly unfulfilled market potential for audio-only games intended for smartphone users. Furthermore, as most of the popular audio-only games are created in the first-person perspective, it also seems important to examine whether or not it would actually be possible to navigate in a game environment solely by using 3D-audio. This could possibly help in creating more exciting audio games, where the player no longer needs to rely on their own imagination of spatial audio, but instead is able to actually perceive the audio as being spatial.

In this article, a framework is presented on how to design navigational 3D-audio games, in which all navigation is based on perceiving the 3D-audio, as opposed to using sonification, auditory icons, earcons or other navigational aids [7, 3, 21], all of which may be added later on in the development process. In this proposed rethink of the design process of 3D-audio games, the focus is shifted from creating complex auditory environments in
which 3D-audio is used as a complementary spatial effect, to creating a functional 3D-audio navigation being complemented by carefully enriching the auditory environment. By following this design principle, games may be made to respond more rapidly and accurately to the player’s navigational input, as well as give the player a sense of full control and feeling of immediate response from the auditory environment, which could help bridge the gap between the hasty action-filled visual games and the slower tempo audio games we have today. In order to test the framework, a game named Sound Hunter was developed, in which all navigation is based on perceiving sounds through synthesized head-related transfer function (HRTF) filtering. The filter used in Sound Hunter is the Pure Data external [earplug~], created by Pei Xiang, David Camargo, and Miller Puckette [24], featuring 368 (or 722 if mirrored to each ear) impulse responses, covering a spherical surface with an elevation of -40 to 90 degrees and an azimuth of 0 to 360 degrees, all of which are interpolated using linear interpolation. The generalized HRTF’s are gathered from KEMAR (dummy head) data sets [10], and the amount of impulse responses used in [earplug~] was considered more than acceptable, seeing as both listening tests and error analyses have shown that 128 impulse responses give satisfactory localization abilities in the azimuth plane [22]. The distance simulation was programmed separately in Pure Data, with the logarithmic loudness function \( L = \log \left( \frac{1}{r} \right) \), where \( r \) is the maximum distance of 8m. Sound Hunter currently features 20 different levels (audio loops and synthesized sounds) varying in difficulty, where the player’s objective is to capture each sound as quickly as possible, by using the iPhone to control player movement. As suggested in the framework, the game was developed in an iterative process, together with both older and younger visually impaired people.

II. Background

Audio-based games are similar to video-based games, with the exception that they are played and perceived through sound and acoustics only [21]. Audio games have many advantages making them interesting for gameplay experimentation. For example, they allow an increased degree of spatial freedom, as no screens are necessary. Furthermore, the computational complexity is usually lower, meaning that less hardware is necessary, making them suitable for portable devices and mobile gaming [21]. Another common favourable argument is that they may lead to an increased level of immersion due to the lacking graphical representation, where the player has to rely more on their own imagination, similar to when reading books [21]. Apart from the earlier mentioned positive effects on physical and cognitive development, audio games can also be used to train a person’s hearing and teach the player how to focus more on what they hear [9].

When building an audio-only game, it is important not only to use self-explanatory sounds, but also to establish agreements early on in the game to convey information correctly to the player [7]. These agreements should build upon metaphors and associative patterns to make it easier for the player to get a sense of what information is important in the game (e.g. the difficulty of the challenge ahead, the current success rate, or scores awarded when completing a game task). It is also important to emphasize the difference between various types of auditory information. For graphics, variations in colours, borders, buttons and other types of design principles are used to label and categorize different types of information. For auditory information, different models have been proposed, where one example is the SITREC categorization system suggested for audio game interfaces [7].

It is also important to distinguish auditory information generated by player activity from the information generated by other sources in the game. This player feedback informs the player whether or not the action was registered by the system [7], and has led to the development of three well-established design methods for auditory interfaces. Auditory icons are recognizable sounds, such as voices or confirmatory sounds [11, 12] and earcons are short musical phrases [3], both of which are associated with various types of information to inform the player of their actions [7]. The third design method, sonification, can be seen as sub-part of the auditory icon, and is the process of mapping abstract data to non-speech sound [21].

It is also possible for game sounds to have several layers of information, often when wanting to enhance the complexity or function of the sound. This additional information may be communicated on a meta-level [7]. In Sound Hunter, distance-dependent meta-level communication was conveyed by altering the cut-off frequency in square waves (to create complex motor-like sounds for easier levels), and by altering the frequency in sine waves (to create distance-dependent audibility for more difficult levels).

When considering the sound characteristics, it is also important that each sound is intelligible and distinguishable. Often, the sounds are accompanied by a musical context. The latter of which seldom is emphasized, but rather added as a backdrop in order to create a scene and set the mood [7, 14]. Another aspect of sound characteristics is that of looping sounds, which may reduce the level of realism in the sound, but is often still desirable in audio games, as it gives the player an overview of the game space [7].

In many action-based audio games, the success of the player depends on interaction based on precise timing [1]. Examples of audio-only action games including 3D-audio are most commonly found in the first-person shooter genre, for example Shades of Doom [13], Terraformers [18], and Demor [4], all of which are made for the PC. Another popular audio game genre is that of adventure games, or exploration games, where three key features are combined: an interesting scenario, the exploration of new worlds, as well as activities of riddle solving [1]. Examples of audio-only exploration games including 3D-audio are Blindsie [6], and Escape The House: A 3D Sound Experience [19], both of which are available for iOS users.

Röber and Masuch [21] attempted to prototype various game ideas for audio-based gaming by using individualized HRTF’s, head-tracking, and a joystick or keyboard for player movement. They created three action-based games (The Frogger Game Remake, Moquitoes, and MatrixShot) and one exploration game (The hidden Secret). Röber and Masuch pointed out several future improvements, such as extending their framework with more advanced sonification and interaction techniques, as well as developing a truly mobile solution, allowing the player not to be bound by webcams, head-tracking devices and other stationary equipment. These issues are also present in the more mainstream audio game Demor [4, 21].

In his Master’s degree project, Graeme [14] developed a 3D-based audio-only game called Blind Fear, featuring advanced
auditory environments. However, similar to the previous studies mentioned here, Graeme found that the more effort he put into creating auditory environments being as advanced and rich as possible, the more difficult it became to actually manage to navigate in the auditory world. During various parts of the development process, Graeme managed to play through the entire game from beginning to end using only audio, but only after simplifying it by sacrificing additional sounds, thus compromising the aesthetical effort [14]. This example not only shows the difficulties that may arise when trying to balance functionality and aesthetics in an audio game, but also the need of including the intended users as early on as possible in the game’s development process to minimize compromises between usability and aesthetics.

One of the more promising games is Blindside [6], requiring only a smartphone and headphones, where only a few sounds are focused on at a time, making the 3D-audio easier to perceive. However, there are still problems with this game. For example, it is quite static and slow, just as many other audio games (e.g. press forward button to move in the game at a pre-determined speed). Blindside also relies on the iPhone’s gyroscope (where the player needs to stand up and spin around), which on the one hand gives more spatial freedom, but on the other hand makes it difficult to play when sitting up or lying down.

For the more mainstream 3D-audio-only games not relying on stationary equipment (e.g. Shades of Doom [13], Terraformers [18], or Escape The House: A 3D Sound Experience [19]), the actual 3D-effects are very difficult to perceive. The games give the impression to have been developed as if they were intended for stereo usage, with some additional binaural sound introduced to the mix. As there are many tools and programming libraries available for creating 3D-audio environments, such as FMOD’s head related transfer function [14], and it can be expected that developers and sound designers decide to enhance stereo environments with various binaural sounds, rather than relying on only stereo.

III. PROPOSED FRAMEWORK: RETHINKING THE DESIGN OF 3D-AUDIO GAMES

All of the above mentioned audio-only games based on 3D-audio claim to present the player with exciting 3-dimensional sound-environments, either through surround sound systems, or by 3D-audio through headphones. However, the focus is usually not on the player’s ability to perceive the 3D-audio in order to use this cue as the main tool for navigation, but rather on enhancing the perceived quality of the stereo environment in order to make it richer and more life-like. For the games focusing more on the navigational purposes of the 3D-audio, the problem instead seems to be their reliance on head-tracking devices or other stationary equipment [4, 21], making them highly inappropriate for relaxed gaming (e.g. lying down as opposed to walking around), or mobile gaming with only a smartphone and earphones.

In order to truly make use of 3D-audio for the purpose of navigation, however, it becomes very important not only to create the 3D-audio environments themselves, but to also understand our natural abilities and limitations when it comes to perceiving spatial sounds as human beings. For example, when filtering a sound using an HRTF filter in order to place the sound in a certain position in a 3D space (e.g. in front of the listener and slightly to the right), the sound will not automatically be perceived as being in front and slightly to the right of the listener without involving movement, either by moving the player’s (head) position, or the sound source’s position, creating dynamic localization cues. The effect has been named the cone of confusion, and arises when sounds in different positions in a vertical circle around either side of the listener’s head have equal inter-aural time differences (ITD’s), and the confusion is eliminated when moving the head (i.e. causing the sound’s relative position to the listener to change, thus altering the ITD’s) [17].

Furthermore, additional stereo sounds being used in audio games may disturb the player’s ability to perceive the HRTF filtering. This is because our ability to perceive a sound as being spatial depends on the above mentioned ITD’s, as well as interaural intensity differences (IID’s), and further qualities of the sound, such as spectral differences, or room qualities simulated by reverberation, all of which may be masked, interfered, or in other ways become inaudible in either ear, causing severe inability to localize the sound object [14, 16]. As many audio-only games using 3D-audio can have over 30 different tracks playing simultaneously [13, 14], where some are in stereo (e.g. game music, instructions, auditory icons or earcons, usually at a relatively high volume), and some may be HRTF filtered, it is therefore not surprising that the 3D-effect becomes difficult to perceive.

There is usually a trade-off whether to use more sounds to enrich the auditory environment, or to use fewer sounds to aid navigation [7, 14]. Most common, however, is to add more sounds, such as auditory icons and earcons, in order to aid navigation through the already over-complex auditory environments [21, 7]. The framework for Sound Hunter was therefore developed in a way ensuring that perception-based 3D-audio navigation was the most important aspect of the game, with further sounds being added only later on in the development process (see Figure 1).

Finally, there exist, to the best of the author’s knowledge, no examples of audio games where it is clearly stated that the game
was developed in an iterative process together with visually impaired people (or other intended user groups for that matter). As modern technology-oriented development processes based on Human Computer Interaction (HCI) strongly suggest that the intended users should be part of even the earliest Lo-Fi prototypes (e.g. paper prototypes, workshops) [20], the including of the users is therefore suggested as one of the key aspects of importance in developing a highly functional and entertaining audio game, being effective, efficient, and also satisfying its intended user group.

The framework proposed to ensure a more successful design of 3D-audio games can be viewed in Figure 1. Especially important are the recurring evaluations of in-game navigation, used to update the sound design aspects. These iterations should be made with the target user group.

IV. METHOD

Sound Hunter’s game development process consisted of four parts involving 10 participants: 1) a pre-study in which other 3D-audio games were evaluated in order to create a framework for developing navigational 3D-based audio games and come up with the initial game idea, 2) three focus group sessions with two visually impaired participants, 3) usability testing with eight (new) participants: four visually impaired, and four sighted participants, and lastly 4) a final evaluation of the game with a participant from the usability tests; a visually impaired commercial audio-only game developer, also highly knowledgeable in the field of HRTF synthesis. Due to the great difficulties in acquiring visually impaired participants for the development of an audio game, the evaluation methods were mainly qualitative. However, some quantitative data-collection was also acquired during the usability tests to determine the game level order by difficulty, as well as to find trends in how the players gaming abilities improved over time. The initial game idea for Sound Hunter is presented first, followed by the adjustments having been made to the game throughout the development process.

V. SOUND HUNTER: THE ORIGINAL GAME IDEA

The original game idea for Sound Hunter was relatively simple. The intention was to create a single-player arcade-based game in which a sound is placed at the maximum distance (8m) and random location in a 3D-space, utilizing the azimuth angle. The player’s objective is to capture the sound as quickly as possible by using 3D-audio as the only navigational aid. All player movement is controlled by the iPhone’s accelerometer data, where leaning the iPhone forward or backwards (the accelerometer’s y-data) leads to the player moving forward or backwards in the game, and leaning the iPhone left or right (the accelerometer’s x-data) corresponds to the player’s head turning either to the left or to the right. Once the sound is reached, a confirmatory auditory icon is heard, and the next level begins, where the levels become progressively more difficult. The game is programmed in Pure Data.¹

VI. FOCUS GROUP SESSIONS 1–3

The participants were the same for all three focus group sessions.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Sex</th>
<th>Blindness</th>
<th>Age</th>
<th>Smartphone</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>M</td>
<td>Born blind</td>
<td>31</td>
<td>Yes, iPhone</td>
</tr>
<tr>
<td>P2</td>
<td>M</td>
<td>Legally blind</td>
<td>57</td>
<td>No, Nokia</td>
</tr>
</tbody>
</table>

P1 had played the occasional audio-based game, but never in depth. He also worked as a usability expert for utilities intended for blind people. P2 had played simple arcade-based games in his youth, but had no experience with audio-only games.

A. Focus group session 1

The first session began by presenting HRTF technology and its capabilities. In order not to bias the participants, an open discussion was held related to the types of games that might be created by using synthesized HRTF filtering, after which the initial game idea for Sound Hunter was presented and discussed. This allowed the gathering of general data and data being specific to the initial game idea.

Results: After some open discussion, both participants suggested that basically anything using this technology that is fun, entertaining and functioning has the potential of becoming a great success. P2 also pointed out that more simple games could be fun, such as the old-school arcade games he had played in his youth, in which the player reaches higher levels under various constraints (e.g. time, enemies). When presenting the game idea for Sound Hunter, both participants regarded the navigational aspect of being able to locate and control sounds in a 3D-based sound-environment as exciting (e.g. being able to hear sounds approach or appear behind the player) – “We are a bit like virgins in this sense. Everybody else has played all of these cool games, but we don’t have this experience, so I think most games using this technology could be fun” (P2). An important conclusion from this focus group session was that all non in-game sounds (e.g. instructions, highscore) can be displayed as text on the iPhone instead of being pre-recorded, as this allows the user to control the language and VoiceOver reading speed as they prefer. Furthermore, we concluded that 20 game levels would be enough for further tests, which could be altered in difficulty by:

- Varying the level of realism in the sound (realism as in its connection to a real-life scenario e.g. crying baby vs. growling tiger),
- Inserting different amounts of temporal silence in the sound,
- Varying the complexity in the sound (e.g. dynamic drum-set vs. sine wave),
- Varying frequency (e.g. different ranges or by approaching our perceptual limitations), and
- Using additional cues related to distance changes being communicated on a meta-level (e.g. varying the level of activity or frequency in a sound object depending on its distance).

B. Focus group session 2

The intention with the second focus group session was to let the users test the navigation to see if it was intuitive, and to get an

¹ Pure data, http://pure-data.info. A planned development is to use libPd to have the application running on a mobile device.
understanding of the learning curve for blind users using the iPhone’s accelerometer to manipulate sounds in a 3D space.

Results: First, the participants’ passive azimuth localization abilities were tested to see if they experienced front-back confusion. For simultaneous listening two pairs of in-ear earphones were used (Pioneer-CL711-G, frequency range 8Hz-22kHz). The sound was positioned and moved forward, backwards and around the listener using a graphical interface. The participants were told to point in the direction of the sound source when being asked of its location. This was done ten times unless an error was made. Both participants pointed in the correct direction all ten times, and expressed that they had a clear perception of the position at all times. Following this, the iPhone navigation was explained and practiced for about one minute. The sound was then repositioned at the maximum distance and random azimuth location, and the participants were asked to capture the sound as quickly as possible. The sound could only be captured within a range of 2% of the total radius, or 0.32m of 16m in width (i.e. the sound had to be almost exactly straight in front or behind the listener in order to be captured in a straight run). After some open discussion, the following conclusions were made:

<table>
<thead>
<tr>
<th>Improvement</th>
<th>Reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The iPhone should be held horizontally, using two hands – “It takes a while to get used to, but it is the most logical way of working with it” (P1).</td>
<td>1. Promotes easier hand movement, and the VoiceOver function alerts the user of the holding position already in the game menu, making it a more intuitive game control.</td>
</tr>
<tr>
<td>2. There should be a training option in the menu. – “It takes a while so if you practice first you’ll get a confidence boost” (P2)</td>
<td>2. Allows the user to train their navigation during an unlimited period of time with a complex and easily heard sound.</td>
</tr>
<tr>
<td>3. The sensitivity should be increased, and the sound’s movement should accelerate when leaning the iPhone more.</td>
<td>3. Promotes easier captures at close distances. – “When it’s close you want to be able to get it fast and not wait for it to spin” (P2)</td>
</tr>
<tr>
<td>4. The game levels should vary in time, with longer times at the first levels and shorter times as the game progresses.</td>
<td>4. To promote a sense of confidence in the player. – “You want to complete at least some of the first levels to get started” (P2)</td>
</tr>
<tr>
<td>5. The sound should decrease more in intensity at further distances.</td>
<td>5. To give a greater illusion of externalization</td>
</tr>
<tr>
<td>6. Apart from the notification sound alerting the player that the sound has been captured, further auditory icons should be included.</td>
<td>6. Countdown to the next level, new highscore, new time record, game over, and dynamic sonification representing the score amounts for each level.</td>
</tr>
</tbody>
</table>

During the navigation, hand and head movements were observed and task completion time was registered. P2’s hand movements were calm and sequential, while P1’s were more rolling, as if using a joystick. P2 captured the sound within 73 seconds, while P1 needed more than 120 seconds. Both P1 and P2 held their heads still.

C. Focus group session 3

During the third focus group session, the participants tested the re-programmed navigation, as well as the current version of the game, featuring 20 levels varying in difficulty and time.

Procedure and results: When re-testing the navigation, the participants felt that the controls were more intuitive. However, they still had difficulties capturing the sound. The score radius was therefore broadened from 2% to 5%, which led to drastic improvements. When playing the game for the first time, they both came to the fifth level. However, as the placement of the levels was based on initial subjective impression of their independent difficulties, as well as the fact that both P1 and P2 expressed that the level difficulties varied a lot, they were allowed to play some of the later levels in the game (note that these levels had shorter times). When beginning at level 11, they both managed to reach level 13. Interestingly, they both accomplished the same amount of levels in both tests, and the atmosphere was clearly competitive. The main conclusions from this focus group were the following:

<table>
<thead>
<tr>
<th>Improvement</th>
<th>Reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Further usability tests were necessary to determine the correct level order.</td>
<td>1. Mixed level difficulty – “I’m not sure exactly why, but some of the levels were much more difficult” (P1).</td>
</tr>
<tr>
<td>2. The score radius should remain at 5%.</td>
<td>2. “I think 5% is the perfect normal difficulty level” (P1)</td>
</tr>
<tr>
<td>3. The time decrease should be mentioned in the instructions.</td>
<td>3. Increases excitement and immersion – “When I knew this it became much more fun playing” (P2).</td>
</tr>
</tbody>
</table>

VII. USABILITY TESTS

The usability tests were conducted to determine the correct game level order, but also to see how well the participants perceived the HRTF filtering and how they used it interactively. Further reasons were to find out which levels were the most difficult and why they were considered difficult, to see if the participants improved in the game over time, to make observations of the participants movements and interactions, as well as to get qualitative feedback on the game in general.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Sex</th>
<th>Blindness</th>
<th>Age</th>
<th>Smartphone</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>F</td>
<td>Sighted</td>
<td>24</td>
<td>Yes, iPhone</td>
</tr>
<tr>
<td>P2</td>
<td>M</td>
<td>Legally blind</td>
<td>52</td>
<td>Yes, iPhone</td>
</tr>
<tr>
<td>P3</td>
<td>F</td>
<td>Born blind</td>
<td>50</td>
<td>Yes, iPhone</td>
</tr>
<tr>
<td>P4</td>
<td>M</td>
<td>Sighted</td>
<td>16</td>
<td>Yes, other</td>
</tr>
<tr>
<td>P5</td>
<td>M</td>
<td>Sighted</td>
<td>15</td>
<td>Yes, iPhone</td>
</tr>
<tr>
<td>P6</td>
<td>M</td>
<td>Born blind</td>
<td>23</td>
<td>No</td>
</tr>
<tr>
<td>P7</td>
<td>M</td>
<td>Sighted</td>
<td>24</td>
<td>Yes, iPhone</td>
</tr>
<tr>
<td>P8</td>
<td>F</td>
<td>Born blind</td>
<td>18</td>
<td>Yes, iPhone</td>
</tr>
</tbody>
</table>

HRTF/binaural sound experience: P1, P2, P3, P4, P5, P6, P7
Notable HRTF/binaural sound knowledge: P6, P7

Procedure: After explaining how the game worked, the participants’ passive HRTF perception was tested similarly as in focus group 2. The participants then got familiarized with the navigation, and were allowed to practice as long as they wanted.
When they captured the sound, it was replaced to let them capture it again until they felt comfortable enough to start playing the game. Each participant played through the entire game including all 20 levels five times each. Simultaneous listening was conducted throughout the test, and the time restrictions had been removed, thus allowing the participants to play each level as long as they needed in order to complete it. The quantitative data analysis consisted of analysing the amount of tips that had to be given for each level to be completed, as well as the level times retrieved through automatic data-collection, to determine level difficulty, and to see how the participants improved between gaming sessions. The qualitative data analysis consisted of observations and a semi-structured post-game interview, in order to obtain additional information regarding game improvements.

Tips: The tips were categorized into amount, participant, level and session. A tip was given if the participant communicated uncertainty connected to something previously explained (e.g. how the navigation worked), or if 90 seconds had passed without level completion. The given tip was always connected to the navigation (e.g. “Try leaning forward”), regardless of what caused the uncertainty, such as a certain level being more difficult, front/back-confusion, or if the participant could not hear the sound object (being the case for levels 19 and 20, where the frequency only becomes audible at closer distances). Almost all the tips were given in the first gaming session, indicating that it took about one session for the participants to properly understand the controls. The exceptions were the final two levels, where the frequencies were more difficult to perceive for the two older participants (P2 and P3), and the tips were therefore spread out between sessions. P2 had a documented hearing impairment on the right ear, and could barely hear level 19 and 20. The completion times were therefore set to 100s for P2 on these two levels for the quantitative data analysis. Furthermore, P8 had severe difficulties concentrating, and could therefore not finish the test. The results were also too biased by tips and uncompleted levels, and P8 was therefore excluded from the quantitative data analysis. Apart from this, no adjustments had to be made, and the quantitative data analysis was performed with seven participants.

Level difficulty: The new level order was determined by comparing the level completion times using a One way Repeated Measures Anova in SPSS. The within-level difference in level completion time was significant $F(5,322, 180.958)=14.31, p<0.01$, and 29.6 per cent of the total variance could be explained by completion time (using the Greenhouse-Geisser correction, sphericity assumed). As expected, the levels’ difficulties did not match the order in which they had been placed (see Figure 2).

As seen in Figure 2, the easiest and most stable levels, were levels 11 to 13, which were the synthesized levels communicating complex distance-dependent meta-level information (level of activity in a motor-like sound depending on distance), and the most difficult levels (despite the broad confidence intervals caused by audibility differences between age groups) were levels 19 and 20, which communicated inaudible to audible frequency changes in sine waves, also depending on distance (20kHz to 5kHz for level 19, and 20Hz to 150Hz for level 20). It should also be mentioned that ITD’s could not be utilized in level 19 due to the high frequency content, and IID’s could not be utilized in level 20 due to the low frequency content, which was part of the experiment to indicate which of these static localization cues are the most important in virtual interactive gaming experiences.

![Figure 2: The comparable difficulty difference between each level.](image)

**Practice effect**: The level times were also used to analyse the practice effect between gaming sessions, performed similarly as level difficulty, using a One way Repeated Measures Anova in SPSS. The within-subjects effects were significant $F(2,441, 339.331)=11.572, p<0.01$, and 10.6 per cent of the total variance in game completion time could be explained by the different gaming sessions (using the Greenhouse-Geisser correction, sphericity assumed). The differences between gaming session one and two $t(139)=3.90, p<0.01$, one and three $t(139)=4.46, p<0.01$, one and four $t(139)=3.62, p<0.01$, and one and five $t(139)=4.41, p<0.01$ were significant. All other differences and pairwise comparisons between gaming sessions were non-significant (see Figure 3).

**Observations**: To get an understanding of how the participants behaved when playing the game, special attention was paid to their hand and head movements, passive and interactive front/back-confusion, as well as their overall improvement and immersion in the game (see Figure 4). Two
participants (P5 and P8) experienced front/back-confusion in the passive HRTF perception test, where P5 adapted immediately while playing, and P8 only adapted while playing levels 11 to 13 (including additional dynamic meta-level localization cues). All the participants used the navigation to eliminate front/back-confusion while playing the game (i.e. by spinning the sound, simulating the dynamic localization cue of head movement). The participants held their heads still, but their hand movements differed, where some used rolling movements, and some used calm and sequential hand movements. As indicated by the quantitative data analysis, it seemed to take about one gaming session (≈15 min including practice time) to fully understand the navigation. Five participants expressed signs of immersion (e.g. laughter, expressions of disgust for insect levels, or through threatening statements directed to the hunted sound object).

![Participant playing Sound Hunter in their natural home environment.](image)

**Figure 4:** Participant playing Sound Hunter in their natural home environment.

**Post-game interview:** The post-game interview consisted of a semi-structured questionnaire with 51 questions, either being nominal, open, or measured on a 7-point Likert scale. The results confirmed most of the observations and quantitative results. The controls were seen as easily understood, and none of the participants considered front/back-confusion a problem, as long as it could be corrected by navigating. Sounds were perceived as more externalized than lateralized, and the overall sound quality and 3D rendering was seen as extremely good. All the participants, even the blind, stated that they improved in the game as a result of focusing attention to their hearing more than usual, or even training their hearing, rather than remembering the game levels. Five participants were highly immersed in the game, whereas three participants were more in a state of deep concentration (or flow). They all felt happy while playing the game, and stressed as a form of adrenaline rush, except P8, who stated that she was currently being examined for ADHD, and therefore always felt stressed and also had great difficulties concentrating or sitting still. This was clearly noticeable also during the observations. Finally, the game in itself was highly appreciated, mainly due to the 3D-audio experience, and all the participants would recommend Sound Hunter to others:

- “The best part was being able to play a game with motion as a visually impaired person, and I would recommend the game to others, even sighted people, as it shows how important our hearing is” (P2)
- “It felt very interesting. I hardly ever play computer games and I thought it was very fun. I think it is good also to train one’s hearing, and I can imagine that you could use this for all kinds of other things. It would probably be a great game for kids with special needs” (P3)
- “I thought the various sounds as well as the whole experience of the 3D sound in the game was cool and fun” (P7)
- “I’ve never played anything like it, and I would recommend it to others because of this” (P5)
- “The best part was the 3D audio and your fantastic simulation! It was really easy to localize sounds. Other games do not provide this. I have never heard anything like what you have done here, and I have tested everything, and of everything I’ve heard, you have come the closest to providing real 3D audio” (P6)

As the results from the focus group sessions and the usability tests showed that the participants could accurately use the 3D audio to navigate by using simple hand movements, the framework was considered applicable for similar developments.

**VIII. Final Evaluation**

The final evaluation was conducted with P6 from the usability tests, a 23 year-old blind audio-only game developer, also highly knowledgeable in the field of HRTF synthesis (will be referred to as P6 also in this section to avoid confusion). After letting P6 test the finished game with the correct level order, a semi-structured interview was held covering the development process and its results, how Sound Hunter may be optimized, as well as what future developments might be made by following Sound Hunter’s framework.

**Optimizing Sound Hunter:** P6 was first asked whether he thought there was a need for audio-only smartphone games.

- “Yes, absolutely, the amount of games and game developers for iOS increases all the time, so the interest is definitely there. I am completely convinced that you could capture the market there if you intend to develop a fully functioning iPhone game.”
- “I have never heard anything like what you have done here, I have come the closest to providing real 3D audio” (P6)
- “To avoid the unfair difficulties between age groups, and because most normal headphones do not support these ranges” (P6)
- “The best part was the 3D audio and your fantastic simulation!” (P6)
- “Front/back-confusion was greatly reduced in Sound Hunter through complex distance-dependent communication, and we discussed additional ways of accomplishing this, such as to introduce front/back sound bounces, or directional beeps. We also concluded that the frequency ranges towards 20Hz and 20kHz should be tighter – “To avoid the unfair difficulties between age groups, and because most normal headphones do not support these ranges” (P6)
- “I thought the various sounds as well as the whole experience of the 3D sound in the game was cool and fun” (P7)
- “I’ve never played anything like it, and I would recommend it to others because of this” (P5)
- “The best part was the 3D audio and your fantastic simulation! It was really easy to localize sounds. Other games do not provide this. I have never heard anything like what you have done here, and I have tested everything, and of everything I’ve heard, you have come the closest to providing real 3D audio” (P6)

**Future development:** Regarding future development, the game could be split into several unlockable “Sound hunting missions”, each of which would correspond to ways of progressively making the game more difficult (e.g. silence, audibility, meta-level communication) – “This is a classic way of getting people to play...
more” (P6). Also, more sounds could be introduced, as well as simple forms of artificial intelligence (e.g. threatening or shooting objects). However, the most important conclusion was that all future developments should be made with Sound Hunter’s underlying framework in mind, such that navigating is still possible.

IX. CONCLUSIONS
When playing an action-based game, whether it is graphical or auditory, the player wants to be transported into another world and perceive that world as if it were real. Arguably, an increased level of immersion due to increased imagination is not necessarily the reason behind the success of action-based graphical video games, and there is little reason why this should be the case for audio games. Creating a sense of auditory spatial presence can only be accomplished properly by connecting our real-world perceptions with the actions and responses in that of the game. Today, audio games based on 3D-audio are extremely rare, with only a handful being available for smartphone users. For these games, the main problems seem to be the inability to use 3D-audio as the only means of navigation, reliance on stationary equipment or physical movement by the player, as well as insufficient usage of HCI methods in the development processes.

In this article, a framework has been proposed for designing navigational perception-based 3D-audio games without relying on imagination or additional navigational aids, all of which may be added later on in the development process. To test the framework, the game Sound Hunter was developed in an iterative process together with both sighted and visually impaired participants. The results indicate that the proposed framework is a successful guidance tool when developing navigational 3D-audio games, and that the need for navigational 3D-audio smartphone games may be met by introducing proper design processes directed to the user group. The results also indicate that Sound Hunter may be used to train a person’s spatial hearing, where complex distance-dependent meta-level communication in synthesized sounds greatly reduces front/back-confusion. Finally, all of the participants would recommend Sound Hunter to others, due to the fast and accurate interaction made possible by the way in which the 3D-audio was perceived and could be used to navigate.

ACKNOWLEDGMENT
The author is sincerely thankful to Kjetil Falkenberg at the Sound and Music Computing Group (supervisor), Funka Nu, Synskadades Riksförbund, Unga Synskadade Stockholm, and Blastbay Studios.

More in depth information regarding Sound Hunter and its development is available in the equally named master’s thesis listed in the references [2].

REFERENCES