

CREATING A MULTI-DIMENSIONAL COMMUNICATION SPACE TO IMPROVE THE EFFECTIVENESS OF 3-D AUDIO

Paul Sauk
MITRE Corporation
Belcamp, MD

and

Duane Satorius
Harris Corporation
Annapolis Junction, MD

ABSTRACT

As the number of voices increases, the ability of the listener to comprehend the content of any one of the voices decreases. Current technology makes it possible to create a three-dimensional (3-D) binaural sound system using an audio headphone and a digital processor to augment the listener's ability to spatially distinguish and locate individual speakers within a communication network. True binaural spatial audio over headphones can manipulate the sound to emanate from any particular point in the virtual 3-D audio space surrounding the listener. This paper discusses several audio dimensions that can be applied to a communication network to improve the effectiveness of a 3-D binaural system, improve the listener's ability to understand what is being communicated and provide the listener with metadata regarding their situational awareness (SA).

INDEX TERMS

augmented reality, binaural spatial audio, GPS, multi-dimensional communication space, virtual listening space, 3-D spatial sound

I. INTRODUCTION

Much has been written about how to create a 3-D binaural sound system, the mathematics and algorithms to support such a system, and the various applications. An area that needs to be explored more fully is how the listener could easily adjust and customize his virtual sound environment to become more aware of his surroundings. Having the ability to dynamically alter or customize a sound environment provides the listener with additional information that was not previously available. This paper will discuss various environments and advantages for having an adjustable multi-dimensional binaural sound system and the architecture for said system. For example, such a system could be used to locate sound sources, to track audible voices that may be friend or foe, to adapt more easily to a limited visibility environment, to hear sensor warnings in unfamiliar surroundings, and to receive signal intelligence of impending danger. All of this information can be provided through audible 3-D binaural

sounds rather than through visual displays permitting the user to keep his eyes fixed on the local surroundings and his ears listening to relevant metadata regarding his situation.

One of the trends emerging in the information technology revolution is the expansion of information received through hearing. Specifically, binaural spatial audio creates a virtual 3-D sound space around the listener that can be used to determine the relative direction of a sound source [1]. The theory behind binaural spatial audio is not new [2]. If the digitally processed sound waves arriving at each ear are identical to those of sound waves received from a real sound source, the human brain will naturally determine the direction and location of the virtual sound source [3], thereby providing the listener with audible information about his surroundings [4]. The virtual sound environment can be created using digital filters. Since the human brain receives audible signals from each ear, only two digital filters are needed to create this 3-D binaural effect.

To simplify matters, headsets are typically used to eliminate crosstalk between sounds entering each ear and to eliminate unwanted background noise [5], [6]. This allows the listener to become immersed in the virtual environment. For a multi-user system, headphones are necessary to augment a communications network by permitting multiple users to receive personalized binaural spatial information that could span several miles [7].

By incorporating commercially available miniaturized optical gyroscopes into a headset and Global Positioning System (GPS) technology into the listener's radio/computer system [5], additional SA information can be provided to the listener that would otherwise be provided visually or not at all. As technology advances, electrical and optical components are becoming smaller, lighter, and more rugged. Currently, optical gyroscopes are small enough to fit within a headset or other headgear [8], and the computer needed to augment the standard monophonic audio into a 3-D binaural audio is small enough to be placed inside the radio.

An optical gyroscope housed in the listener's headset and

thus fixed to the listener's head is necessary to track the listener's head motion and orientation. This information is fed into the listener's computer to keep the virtual sound sources fixed relative to their real-world physical locations. Hollier et al. describes this scenario as world-stabilized virtual sound [9]. The head-tracking information is combined with other location information received by the listener in the form of metadata attached to the communication signal. To simulate the voices in a 3-D binaural environment around the listener, the computer must produce slightly different auditory localization cues through a digital filter customized for each ear [9]. The sound arriving at each eardrum undergoes multiple changes that provide the listener's brain with cues regarding the location of the sound source. Some of the changes are caused by the human torso, the head, the ear pinna, and the ear canal. Collectively, these changes are mathematically summed up into a Head Related Transfer Function (HRTF). The digital filter used to incorporate the HRTFs converts the 1-D monophonic sound into 3-D binaural sound [2], [5], [9], [10]. Research has been conducted to methodically examine the degree each localization cue has on the brain which allow humans to determine the exact location of the sound source [6], [10], [11]. The result of this research is an approximate digital filter that permits the listener to locate sounds emanating from different spatial locations to an accuracy of ± 12 degrees [12]. Depending on how the computer adjusts the HRTFs, the listener can dynamically move about any sound environment of his choosing. A detailed mathematical explanation has been provided by Faller and Baumgarte along with the results and advantages of a multi-talker environment [2].

This paper will discuss some of the various simulated environments that the listener can create using this binaural audio system to improve his situational awareness along with applications and benefits for each. The environments discussed are selective filtering, selective relevance, alerts and warnings, intelligence infusion, navigation aid, localization enhancements, and telepresence. A high level review of the architecture will then be described followed by methods used to calibrate and customize a binaural audio system.

II. VARIOUS AUDIO DIMENSIONS

A. SELECTIVE FILTERING

As the amount of audio information being received by a listener increases, the relevant information becomes diluted with unwanted background sounds. To reduce the dilution and thereby improve the signal-to-noise ratio, the user can select which sounds are of interest, and thereby

reduce or eliminate those sounds that are not needed. Humans have a limited ability to focus on selected voices even though background voices may be louder. This has been described in various papers as the "cocktail party effect". A binaural audio system supports this natural ability by placing the voices into a 3-D virtual environment.

A type of selective filtering is group selection. Being able to hear where a person (or group of people) is located just by listening to their voice is one of the more commonly used applications of a binaural audio system. For example, a military reconnaissance mission may involve four groups of soldiers with each group going in a different direction to survey the surrounding area. Instead of listening to all the various conversations occurring in the entire communication network, each group could select their own binaural environment. This would limit the voice communication to just those within their own group. So if soldiers of one group were to spread out in a crowded urban environment and lose sight of each other, they would still be aware of each member's relative location belonging to their group. Their binaural audio system would inform everyone in the group of their relative location by simulating their 3-D audio positions around the listener. In addition, everyone in the group would understand the relative audio positions of the others by simply listening to the binaural voices during communication. Thus, the soldiers could keep their eyes focused on their surroundings and their ears listening to relevant data.

B. SELECTIVE RELEVANCE

Another configuration is combining any number of dimensions such as 3-D (binaural), 2-D (stereophonic), 1-D (monophonic). This capability provides the listener with the desired audible information to help improve his understanding of the environment. In certain cases, the listener may not want to eliminate all the unwanted audio information. Rather than keep the less relevant audio in the same audio dimension with their desired binaural sound sources, the listener could change the less relevant audio to a monophonic (1-D) or stereophonic (2-D) dimension. The effect of changing sounds from binaural to monophonic/stereophonic audio places the sound in the listener's head and signifies a different level of relevancy or importance. This removes any localization cues for that audio. Begault calls this effect the Inside Head Localization (IHL) [12]. Additionally, the listener would be able to adjust the decibel level of the 1-D or 2-D audio to the level of importance the listener desires for that dimension.

An example of how the listener might benefit by combining various audio dimensions would be to have the commander from the previous example listen to the voices of the other three commanders, but in monophonic audio. Thus, the commander can distinguish between the binaural voices of soldiers in his group and the monophonic voices of other commanders. Furthermore, having the ability to listen to the other commanders provides the listener with important information regarding his overall situational awareness.

C. ALERTS AND WARNINGS

In addition to voice communication, other types of information can be superimposed into the sound field, for example, sensor information received by a binaural audio system from the surrounding area. These sensors could be any type including biological, nuclear, or chemical and would be broadcasting a coded signal if activated. If chemical sensors have been dispersed over a large area prior to the arrival of the reconnaissance group and the group enters an area when one of the sensors detects a lethal chemical toxin, the sensor would be activated and begin transmitting a signal that would have GPS metadata and a digital code attached to it. When this RF signal is received by the soldier's radio, their binaural computer determines where the activated sensor is relative to the soldier, determines if the soldier is in danger and if so, translates the digital code to provide the soldier with an audio warning in the form of an earcon. An earcon is a computer generated voice signal that represents the digital code. The earcon would translate the received code to a simple phrase, "Chemical toxin detected. Stay away!" The audible warning would thus alert the soldier that he is too close to the lethal toxin and by listening to the binaural message, the soldier would know naturally which direction to move towards safety.

D. INTELLIGENCE INFUSION

As with alerts and warnings, combining intelligence information with the binaural audio environment could significantly augment the combat power of the front-line soldier. In the past, the problem has been too much information coming into a command center and too little relevant information going out to the soldiers on the front line in a timely manner. In an ideal scenario, the intelligence could be broadcasted with relevant GPS data. Based on this GPS position, the soldiers that need the information immediately would receive it via the selective relevance mode. In other situations, intelligence could be broadcasted from the command center to only those soldiers that need it and would be received via the selective filter mode.

For instance, in urban warfare, sensors would be distributed throughout the cities to detect various events. If a group of soldiers were to go out on a rescue mission with their binaural audio system, they could combine two audio environments to augment their situational awareness. Their binaural audio environment previously discussed in Section A would be beneficial if they had to disperse due to an ambush. Every soldier would know where other soldiers from their group were located by simply listening to their binaural voice communications. In combination with that environment, the audio system could combine the intelligence arriving from nearby sensors.

As an example, certain sensors located throughout a city could be designed to detect gunfire. Once a sensor has been activated, it would transmit a signal containing the spatial location of the gunfire along with a specific earcon code for gunfire to all the soldiers. For a soldier in the city surrounded by buildings, his binaural audio system would receive the signal from the sensor. His binaural audio system would then determine if his GPS location is within the effected zone, and if so translate the earcon identifier into binaural audio saying "gunfire detected in this direction". The earcon would keep repeating every few seconds until the soldier deactivates the earcon message. By listening to the binaural earcon, the soldier can locate where the gunfire occurred. In another situation, an explosion would be detected by a sensor and once activate, would transmit the spatial location of the explosion along with a different earcon code to all the soldiers in the city. In both scenarios, the audio intelligence is being provided to the soldiers in real-time to warn of immediate danger in an area, thus the soldiers can keep their eyes on their surroundings while listening to relevant information.

In a final scenario, if a wanted suspect is located at a particular café in a city, intelligence data would be broadcasted to all soldiers but only those soldiers who happen to be located near the café would receive intelligence to aid in capturing the suspect. All the soldiers within a few blocks of the café would hear the earcon saying "Capture me, I'm wanted!" The soldiers would then carefully move in the direction provided by their binaural audio sound system to locate the café and capture the suspect.

E. NAVIGATIONAL AID

The navigational environment would be used by anyone who needs to get from point A to point B in unfamiliar surroundings. For example, if soldiers needed to be extracted from a hostile area, a short signal containing information about the time and location of extraction

would be received by their personal radio/computer. This signal would be combined with the radio's current location using its built-in GPS and translated by an internal computer to provide the soldier with three pieces of audible information. First, the received earcon would provide binaural audio indicating the direction of the extraction point. Next, the earcon would tell the soldier the distance remaining to the extraction point. Finally, the earcon would tell the soldier how much time is left before the helicopter arrives. Thus, the soldiers would hear an earcon repeat, "Extraction point is two miles away. Thirty-two minutes remaining." The earcon is simulating a voice coming from the direction of the extraction point. The internal computer would update the audible information every few seconds, and the soldier's HRTFs would constantly be updated to guide them to the correct location.

This audible navigational environment could be combined with other audible environments to provide the soldier with additional information about his surroundings. For example, the soldier may need to communicate with other soldiers from their group that may not be within line-of-sight but will also be headed toward the same extraction point. Every soldier could hear the approximate position of other soldiers. If a soldier is wounded and is having difficulty walking, his binaural audio communications could guide nearby soldiers to his location to assist him in getting to the extraction point.

F. LOCALIZATION ENHANCEMENTS

Another audio dimension would be to shape the binaural sound field to provide the listener with better localization cues. The normal binaural sound would be combined with a Binaural Room Impulse Response (BRIR) to create a reverberation effect. The BRIR includes information about all acoustical properties of a room, including the position and orientation of the sound source, the listener, the room dimensions, the wall's reflective properties, etc. Thus, depending on the size, shape, and wall material of a room, the sound source located at one end of the room has different sound properties when heard by a listener at the other end of the room. An example of this technology is provided in most sound systems that are purchased today. These systems have several different sound effects to give the listener the feeling of sitting in an auditorium, a stadium, an inside theater, an outside theater, etc. Research has been conducted to demonstrate the improved localization capability derived from BRIR to give the listener the perceived effect of sound bouncing off walls of differently shaped rooms [15]. By utilizing different virtual rooms to represent radial distances, the listener can better estimate how far the speaker is transmitting from his

position. As an example, if a group of soldiers were scattered over a two mile wooded area and their binaural audio systems were set to provide normal binaural sound for distances less than 100 feet, provide a narrow room binaural sound for distances between 101-1000 feet, and a long narrow room binaural sound for distances greater than 1000 feet, all the soldiers would know by listening to the binaural voices the relative direction and approximate distance between them and the speaker.

G. TELEPRESENCE

The telepresence dimension permits the listener to virtually displace himself into any environment to gain a better understanding of the activities occurring in that area. Using the telepresence dimension, combat commanders would be able to more effectively understand the military operations that are occurring on any particular battlefield. Any commander could virtually transport himself to the front line by programming their binaural audio system with the GPS position of any location on the battlefield. For instance, the commander could select any officer on the battlefield and through telepresence, he would be able to hear the voice communications and virtual positions of all his officers relative to the selected officer. This binaural audio would complement the visual information the commander is receiving from unmanned aerial vehicles flying over the battle site. By being virtually immersed into this combat environment, the commander can make better informed decisions.

II. ARCHITECTURE

A high-level block diagram illustrating how a multi-dimensional binaural communication system would be designed is shown in Figure 1. All the hardware components needed to construct the 3-D binaural audio system are currently available and off-the-shelf. Kan et al. discusses their experimental results using their off-the-shelf components to construct a limited spatial-audio communication system. Some of the limitations Kan discusses are 1) no means for providing the computer with the listener's head orientation or motion 2) the speaker's positional update rates were not fast enough and 3) his communication system was limited to one speaker [1]. The architecture shown in Figure 1 has been designed to overcome these limitations by incorporating a three-axis optical gyroscope to determine the listener's head orientation, and the firmware and software are imbedded in a high-speed digital signal processing chip for faster operation to provide the listener with a real-time binaural communication system.

The ability to customize an individual environment or to

combine various audio environments to maximize the overall performance of a binaural communication system resides primarily in the software design of the sound field environment controller shown in Figure 1. Other than the possible conceptual operation of a multi-dimensional audio system, the details of the software algorithms are left out of this paper and will not be discussed.

All the users in the communication network would have a standard radio enhanced with binaural audio capability and a headset. The binaural audio enhancements would not affect the radio's transport layer, thus encryption for security purposes and modulation techniques to prevent jamming would not be altered. The radio would transmit the encrypted, modulated radio frequency (RF) signal that is commonly accomplished with today's military radios. However, attached to the encrypted signal at the application or content layer would be metadata describing the speaker's GPS position and other relevant data. When the signal is received by the listener, the radio performs its normal decryption, demodulation, and decoding of the RF signal. However, decoding the signal would also involve separating the voice and the metadata, which is then sent to the sound field environment controller shown in Figure 1.

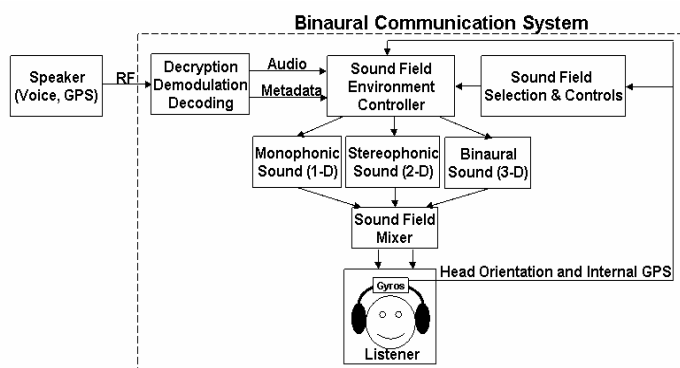


Figure 1. Block diagram of a Binaural Communication System in the receiving mode

All the external data is streamed into a high-speed digital processing chip that is integrated into the radio along with the listener's current head orientation, his selected audio environments and his current GPS location. As shown in Figure 2, the listener would be using a headset that would have a built-in three-axis optical gyroscope. Current technology permits these components to be housed in a relatively small package. The gyroscope would send data to the sound field environment controller regarding the listener's current head orientation and the external speaker's current location. Figure 2 shows the three axes of head rotation and the location of the gyroscope on top of the headset.

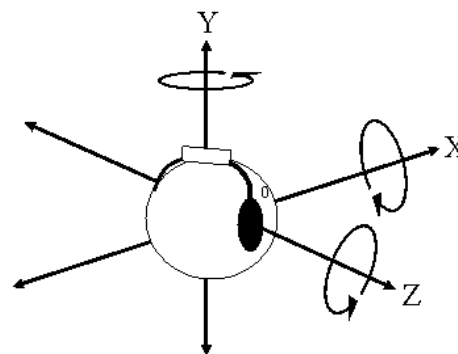


Figure 2. The 3-axes of head rotation: X-roll, Y-yaw, Z-pitch

The sound field environment controller has the ability to either fix the binaural sound relative the listener's head which is called head-fixed, or it could keep the simulated speaker at a virtual position relative to the speaker's actual position regardless of listener's head motion. This second alternative is called world-fixed and both cases are illustrated in Figure 3 below.

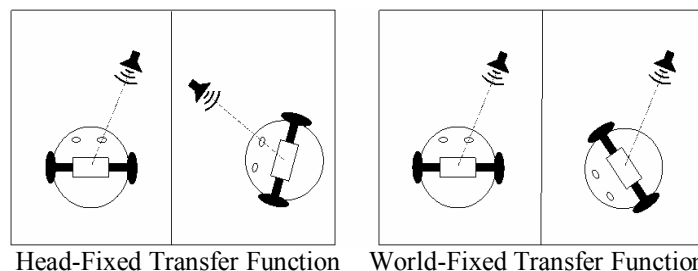


Figure 3. Illustration showing the difference between head-fixed and world-fixed environments

The computer then calculates the virtual sound position of the speaker relative to the position of the listener, determines the head orientation of the listener, makes the adjustments to the left and right Head Related Transfer Functions, and sends the audible voice to the listener's headset. If signals from other speakers are being received simultaneously, the same algorithms mentioned above are applied to those signals and mixed with the voice audio of the first speaker. Kan et al. describes this type of communication system as full-duplex [1]. It should be noted that interference caused at the carrier wave level of the radio will not be changed by the binaural audio system, however interference described as irrelevant audio voices at the application or content level of the radio can be filtered based on the level of importance established by the user. To maximize the performance of the binaural audio system, the user can choose the hierarchy of audio voices and choose the type of audio (monophonic, stereophonic, binaural, stadium, hall-way, etc.) that will be applied to those voice signals. All other undesired audio voices will

be filtered out completely to maximize their situational awareness.

As discussed in this paper, the listener may choose to customize his 3-D audio environment by altering the signals within any environment. Or the listener may decide to combine several audio environments to receive situational awareness information located in his physical vicinity. The number of possible combinations or applications is unlimited. By properly designing the human interface, the user of this binaural communication system could easily select and customize his audio environment to maximize the performance of the binaural audio system.

II. OPERATIONAL IMPLEMENTATION

Because every person has a unique set of head related transfer functions, the user of this binaural communication system would need to perform a quick calibration routine that would be user friendly. This calibration routine will be necessary to customize the decibel and frequency range output of the binaural audio system for both ears of the listener to maximize the performance of the system. By listening to an automated sequence of commands, the listener will calibrate his binaural computer by using a combination of voice commands. When the system is turned on, an earcon will tell the listener to say "Yes" when he hears the simulated sound coming from as far away as possible to maximize the accuracy of the binaural system. This step will cycle through several possible HRTFs to allow the listener to select the most distant virtual location to properly calibrate the system. In the next step, the earcon would tell the listener to face the direction of its simulated voice and say "Yes". This step will calibrate the three gyroscopes in the headset so that the computer can record which direction the listener's head is pointed for the selected HRTFs. Finally, a sequence of earcons would describe each environment to the listener. Each earcon would be spatially separated from the other earcons. By pointing his head in the direction of the earcon and saying "Yes", the listener will select their desired environment. If the listener wants a combination of environments, he would simply point his head in the direction of any additional environment and say "Yes". A similar scenario could be used to adjust or customize the conditions within any particular environment. Once the 3-D binaural audio system has been set properly, the listener would say, "Run".

Another scenario for setting up a group of binaural radios simultaneously would be to wirelessly program the desired set of radios that may be used by a group of individuals on a particular mission. A software application running on a standard computer could be used to allow the user to select

which radios are to be programmed, the audio environments that are to be combined, how each audio environment is to be customized for that mission, and which radios are specifically customized to transmit only in monophonic sound, but receive in binaural sound. Once the set-up procedure is finished, the user would simply transmit the program to the other selected radios. Each individual user would then need to calibrate their headset.

III. CONCLUSION

The compactness and speed of today's technology is sufficient to produce a multi-dimensional binaural communication system. At the highest level, a binaural communication system as described in this paper would enhance the listener's situational awareness, reduce the listener's need to take his eyes off his surroundings, and provide the listener with audible information that may not or could not be received visually. In this paper, we proposed several different 3-D audio environments, various applications that would improve the effectiveness of a binaural audio system, a high-level architecture for the system, and a possible scenario for calibrating and customizing the system. Future work will focus on understanding and improving the HRTFs to improve the 3-D spatial audio accuracy, to develop the software code to implement the various environments in real-time, to program the digital signal processing chip with the proper algorithms, and to construct the hardware prototype to test the system in the field.

From urban warfare in unfamiliar surroundings to firefighting in smoke-filled rooms, the need for such an advanced audio communication systems to augment situational awareness has never been greater.

ACKNOWLEDGMENT

The authors would like to thank Harris Corporation for allowing them the time to investigate this idea and time to write this paper for publication.

REFERENCES

- [1] A. Kan, G Pope, C. Jin, A. van Schaik, "Mobile spatial audio communication system," *Proceedings from the International Conference on Auditory Display*, Sydney, Australia, 2004.
- [2] C. Faller, F. Baumgarte, "Efficient representation of spatial audio using perceptual parameterization," *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA '01)*, New Paltz, NY, 2001.

- [3] H. Moller, "Fundamentals of binaural technology," *Applied Acoustics*, vol. 36, pp. 171-218, 1992.
- [4] D. Keating, "The generation of virtual acoustic environments for blind people," *Proc. 1st Euro. Conf. Disability, Virtual Reality & Assoc. Tech.*, Maidenhead, UK, pp. 201-207, 1996.
- [5] M. Karjalainen, M. Tikander, A. Harma, "Head-tracking and subject positioning using binaural headset microphones and common modulation anchor sources," *International Conf. on Acoustics, Speech, and Signal Processing*, vol. IV, pp. 101-104, 2004.
- [6] M. Tikander, "Sound quality of an augmented reality audio headset," *Proc. of the 8th Int. Conf. on Digital Audio Effects (DAFx '05)*, Madrid, Spain, 2005.
- [7] M. Noisternig, T. Musil, A. Sontacchi, R. Holdrich, "A 3D real time rendering engine for binaural sound reproduction," *Proc. of the 2003 International Conf. on Auditory Display*, pp. 107-110, 2003.
- [8] K. Suzuki, K Takiguchi, K Hotate, "Monolithically integrated resonator microoptic gyro on silica planar lightwave circuit," *Journal of Lightwave Technology*, vol. 18, no. 1, pp. 66-72, 2000.
- [9] M. Hollier, A. Rimell, D. Burraston, "Spatial audio technology for telepresence," *BT Technology Journal*, vol. 15, no. 4, pp. 33-41, 1997.
- [10] B. Cunningham, "Applications of virtual auditory displays," *Proc. of the 20th Annual International Conf. of the IEEE Engineering in Medicine and Biology Society*, vol. 20, no. 3, pp. 1105-1108, 1998.
- [11] M. Gerzon, "Surround Sound Psychoacoustics," *Wireless World*, vol. 80, pp. 483-486, 1974.
- [12] D. Begault, "3-D sound for virtual reality and multimedia," Moffett Field, CA: Ames Research Center, pp. 53-54, 2000.
- [13] D. Zotkin, R. Duraiswami, L. Davis, "Creation of virtual auditory spaces," accepted *IEEE Trans. Multimedia*, 2003.
- [14] R. Duraiswami, D Zotkin, N. Gumerov, "Interpolation and range extrapolation of HRTFs," *International Conf. on Acoustics, Speech, and Signal Processing*, vol. IV, pp. 45-48, 2004.
- [15] I. Biedermann, O. Jorgensen, M. Matejka, S. Merchel, D. Poncet, M. Sorensen, "Reproduction of Surround Sound in Headphones," *Faculty of Engineering and Science, Aalborg University*, Group 960, 2004.