A Systematic Review of the Methods and Experiments Aimed to Reduce Front-Back Confusions in the Free-Field and Virtual Auditory Environments

Oana Balan
University POLITEHNICA of Bucharest, Faculty of Automatic Control and Computers
Splaiul Independentei, 313, Bucharest, Romania
oana.balan@cs.pub.ro

Alin Moldoveanu
University POLITEHNICA of Bucharest, Faculty of Automatic Control and Computers
Splaiul Independentei, 313, Bucharest, Romania
alin.moldoveanu@cs.pub.ro

Florica Moldoveanu
University POLITEHNICA of Bucharest, Faculty of Automatic Control and Computers
Splaiul Independentei, 313, Bucharest, Romania
florica.moldoveanu@cs.pub.ro

ABSTRACT
Human spatial representation is determined by the interaction of a wide range of stimuli, including visual, neuromotor and acoustic information. In a virtual acoustic environment that intends to simulate the perception of sound as in real-world conditions by rendering the audio stimuli over headphones, the accuracy of audio presentation is of the utmost importance. Nonetheless, a well-known problem which affects binaural audio localization is represented by the front-back confusion, a situation in which the listener perceives the sounds which come from the front as coming from the back and vice-versa. Over the years, many theoretical and practical approaches have been devised in order to reduce the incidence of front-back confusions, including head movement, source movement, sound filtering using early reflections, the simulation of reverberant environments or anthropometric estimation. This paper aims to study, review and compare the most relevant methods and experiments designed to decrease the rate of front-back confusion errors that appear in the case of synthesized 3D sound delivered over headphones in virtual auditory displays and in real-world conditions.

Author Keywords
Spatial hearing; sound localization; virtual auditory environment; front-back confusion.

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms
Human Factors; Design; Measurement.

INTRODUCTION

Spatial hearing is an essential prerequisite that influences the human perception and behavior in complex communication environments [1]. Interacting with the perceptual elements of the extra-personal field, determining their location, direction and distance represents a fundamental component of our everyday life. Additionally, the hearing sense can cope with stimuli located outside the field of vision, all around the 3-dimensional space. What is more, auditory cues raise awareness about the surroundings, lead visual attention and transmit a wide range of sensorial information.

The audio spectrum incorporates key information and features that define the entire auditory perception of the environment- relevance and familiarity to the user, direction (in both azimuth and elevation) and distance cues. The auditory perception transforms the signals received at the eardrums into significant information at the brain level, creating a general spatial representation of the environment.

A virtual auditory environment aims to deliver to the listener the same hearing sensations that he would have in real-world circumstances. In order to be effective, it should convey a high degree of sound accuracy (in respect with the direction of sound, distance and acoustic conditions- reverberation, reflections, environmental spaciousness), a high level of realism and an acute sense of presence and immersion. In a virtual auditory environment in which the audio information is rendered over headphones, the listener's ability to locate the sources of sound is lower than in real-world conditions. This is due to the problems that arise during binaural reproduction and simulation with non-individualized (Head Related Transfer Function), the lack of head dynamics (in real-life listening, head rotations are a natural approach towards audio disambiguation) that conduct to a decreased localization performance and to an increase in the front-back confusions rate.

Front-back and back-front confusions are common reversal errors which consist in a series of discrimination ambiguities for the sound sources located in the frontal and in the rear hemifield. As a result, the listener perceives the sources located in the front as originating from the back and vice-versa. Various methods have been designed in order to reduce the incidence of front-back confusions, ranging from the simulation of head rotations, virtual sound image displacements, the use of individualized HRTFs, the addition of early reflections, the reproduction of reverberant environments or HRTF estimations based on anthropometric measurements.

This paper aims to present, describe and compare the most notable approaches (methods, techniques and experimental procedures) that have been devised over the years in order to decrease the rate of reversal errors under normal listening conditions (when the sound is conveyed by loudspeakers) and virtual sound source conditions (when the binaural synthesized audio stimuli are rendered over headphones).

SOUND LOCALIZATION

Sound localization is an essential aspect of our everyday lives, as it is highly related to our performance and personal security. Humans are able to localize sounds from the very beginning of their lives and localize with great accuracy even in unfavorable conditions [2]. For instance, the direction and distance of an emitting sound can be perceived earlier and more accurately using the hearing sense than by searching the source into the visual field. As hearing is a 3-dimensional sense, it is based on the association of 2 distinctive localization judgments: the auditory perception from the horizontal plane and the sensory awareness that comes from the vertical plane.

Sound localization depends on various physical, psychological and behavioral factors that influence the auditory system. In the field of psychoacoustics, they are known as localization cues. The main localization cues which affect spatial perception are: monaural, binaural, dynamic, visual and memory cues. While the monaural cues are related to the degree of filtering of the external ear, head and torso, binaural cues refer to the difference in time and intensity level of the audio signals reaching the inner ear. In addition to this, the anatomical characteristics of the listener's body influence sound localization by creating an individualized feature that uniquely modifies the frequency spectrum of each incoming sound. This specific acoustic component is known as the Head Related Transfer Function, a response that describes how the ears perceive an audio signal coming from a particular point in space. Besides, sound localization depends on the head or body displacements, the interest and motivation of the listener, a certain level of training and adaptation, the familiarity with the sound or the interference with the visual sense.

LOCALIZATION ERRORS

Reversal errors represent localization judgments indicating to the opposite direction than the actual position of the sound source. The listener makes an ambiguous supposition concerning the location of the incoming audio signal, pointing to its mirror image across the interaural axis - they can be front-back, left-right or up-down confusions. For instance, for a sound source located at 60 degrees in the frontal hemisphere, the listener can perceive an apparent position at 120 degrees, in the rear hemisphere [10].

Reversal errors are caused by the spherical shape of the head, by the reflections of the environment, interferences with other objects or sounds and modifications in the spectrum of the wave, caused by identical levels of ITD (Interaural Time Difference) and ILD (Interaural Level Difference), especially under headphones conditions playback. For any azimuth angles in the frontal hemifield, between 0 and 90 degrees, the locations which share the same ITD are symmetrical to the interaural axis, in the range 90-180 degrees. ITD and ILD are ambiguous localization cues, as they share a single value not only to a specific position in space, but to a larger area of sound sources, called "cone of confusion" [3]. Assuming that we have a fixed, spherical head, with both ears situated symmetrically on the left and right side of the head, the multitude of spatial locations that share the same values of ITD and ILD describe the surface of a cone which contains an infinite number of ambiguous points that are difficult to be identified and localized [4]. Front-back confusions appear more commonly in the presence of short, discontinuous or narrowband sounds. While left-right (or right-left) confusions are rather rare (less than 2% of the total localization errors), front-back confusions are more often encountered, especially in virtual auditory environments rendered over headphones.

The ITD is determined by the path length difference between the sound source and the two ears of the listener. The iso-ITD represents the hyperbolic surface of rotation that is symmetrical to the interaural axis. The intersection of the ILD and ITD volumes is a torus-shaped volume, called torus of confusion, whose volume increases as the sound source approaches the median plane [5].

Virtual auditory displays use 3D sounds to convey to the listener the same sound perception as in real-world environments. They enhance immersion and give the feeling of actually "being there". Spatial auditory displays convey spatial information concerning the constitutive elements of the environment to the listener by using the hearing sense as the main perceptive channel. Thus, the locations are encoded by directional sound sources which give the user the ability to investigate and to navigate the virtual settings as he would have explored a real-world, natural setting. Headphone playback is preferable because it provides full control over the audio representation of the environment at both ears, in a natural and realistic style. In addition to this, real-time audio rendering, in combination with head-tracking devices are able to simulate the change in audio perception which accompanies the listener's head movements [4]. Both accurate localization perception and increased realism and immersion are enhanced by 3D binaural sounds. Thus, they need to be reproduced as faithfully as possible, in order to provide improved naturalistic hearing experiences for the listeners who are engaged in the virtual auditory environment. The most employed techniques aimed to enhance the sense of presence and to offer an accurate audio spatial perception are: technological and methodological approaches designed to maintain constant the listener's orientation towards the sound source by using a head-tracking device or by employing a virtual head-tracking algorithm, individualized pairs of HRTFs that correspond to the listener's anatomical features- ears, shoulders and torso and audio processing techniques for the synthesis of reverberant environments, based on HRTF filtering with early sound reflections which are supposed to play an important role in sound localization, as the listener makes his localization directional judgment based on the first audio waves arriving at the ears. Nonetheless, a virtual auditory display will not reach its purpose of conveying a full auditory representation of the environment unless it will not offer an accurate sound localization perception.

Headphone rendering of the sound in these types of applications leads to a series of limitations that cause localization errors and perceptual misjudgments. For instance, the first concern is caused by the inability to correctly localize the sound sources coming from the front and from the back hemifield. The second is that sound localization cues are highly individualized and user-specific. To ensure accurate sound localization, measurements of the HRTF for each listener are highly required. Thirdly, in order to convey a relevant spatial and directional perception, the spectrum of the sound needs to have a broad bandwidth, composed of a large variety of frequency ranges and should incorporate monaural cues pointing to the incoming spatial location in both the horizontal and the vertical plane [6].

FRONT-BACK AND BACK-FRONT CONFUSIONS

Different experiments showed that the incidence of front-back confusions is higher for synthetic spatial sources and 3D audio rendered over headphones in virtual auditory environments. This is caused by the use of non-individualized HRTFs and by the quality of the playback technique. Some authors consider that front-back confusions are not localization errors. However, the majority of researchers believe the contrary, treating and evaluating the reversal errors separately from the angular precision errors that occur out of the cone of confusion.

HRTFs are highly dependent on the anatomical dimensions and characteristics of the human body - the internal and external ear, head, shoulders and torso, which significantly interfere with the auditory perception. Front-back confusions tend to be more often encountered than back-front confusions, especially in the horizontal plane, along 0 degrees in azimuth.

The experiment of Wenzel et al. [7] showed that the use of non-individualized HRTFs led to a significant increase in the rate of reversal errors for the virtual source condition. This is due also to the degradation of spectral cues during HRTF synthesis. The small errors added by the HRTF processing and the use of generic HRTFs introduce jitter and noise, conducting to spectral modification in the shape of the impulse response (shifts and reductions of the peaks and valleys specific to the HRTF). In addition to this, it has been observed that visual stimuli play an important role in sound localization for the sources situated on the cone of confusion.

Reversal errors can be disambiguated by introducing visual stimuli, dynamic cues (such as head or source image movement), the use of familiar sounds and the simulation of reverberant environments enhanced by room acoustics features. In what concerns head rotations, the experimenter needs to use a head-tracking device to measure the directional motion performed by the listener. Nonetheless, head-trackers are expensive, impractical, difficult to calibrate and to use in real-world environments.

Wightman and Kistler [8] [9] showed that front-back ambiguities can be reduced if the user is aware of the location of the sound source. Thus, by managing the motion of the incoming audio signal relative to his own frame of reference, the listener obtains similar front-back localization accuracy and reversal ambiguity decreases as in the case of real or virtual head movements. As during free-field simulation it is difficult to move the sound sources and to have control over them, this form of front-back confusion resolution is available only in 3D audio-based virtual auditory environments. In addition to this, several researchers have intended to improve generic HRTFs through anthropometric customization based on the anatomical measurements and shape of the listener's external ear, as well as on acoustic raytracing [10].

METHODS TO REDUCE FRONT-BACK CONFUSIONS

Head movement experiments

In 1940, Wallach [11] demonstrated that head movement during the presentation of an audio signal represents a reliable cue for the differentiation of sound direction, providing a complete localization perception and spatial judgment. During head movement, the level of the binaural cues is modified and the pattern of their change determines the sound direction.

The angle between the direction of sound and the interaural axis is called "lateral angle". During head movement, the position of the interaural axis is being modified, causing a displacement of the lateral angle. It is this shift of orientation which leads to a modified perception of the incoming sound source. Wallach noticed that not only a simple head rotation is effective for accurate localization, but also a tilting from side to side (or nodding). He concluded that any head movement is able to improve sound localization as long as it produces a rotation around an axis contained in the median plane of the head.

The perceptual factors which need to be taken into account in the process of localization are the modifications in the values of the binaural cues and the data sets representing the tracked positions of the head during actual movements. During head movement, there are 3 categories of sensory data that affect the localization perception: proprioceptive sensations obtained from the neck muscles involved in the process of movement, visual information and the stimulation of the vestibular apparatus.

Body movement experiments

The study described in [12] presents a method that uses body motions (not head movements) for mitigating front-back confusions that occur in a virtual auditory environment. The initial azimuths of the sound sources were 40, 60, 80, 100, 120 and 140 degrees to the left and right. The tracks the listener had to pursue resulted in final changes of 4, 8, 12 and 16 degrees in azimuth between the initial and the final position. The perception of front and back sound source locations (48 spatialized continuous noise bursts, stationary in respect with the real-world frame of reference) was recorded before and after the user walked the distance indicated. The results demonstrate an improvement in the ability to localize front-back audio sources as a result of walking in the forward direction, alongside the source. This is due to the continuous change in the azimuth and distance values that makes the listener to be aware of his position in space and to continuously update the perception of the sound location in the front or in the rear hemifield. The azimuth differences (between the initial and the final position) of 12 or 16 degrees offer the best rate of front-back accuracy. As a result, the body dynamic cues are able to reduce the incidence of reversal errors and to disambiguate localization on the cone of confusion within a relatively narrow angular range.

Source movement experiments

In the sound localization experiments in which head movements are allowed, a significant drawback lies in the difficulty of using special hardware systems, such as head trackers. Yet, the sound image movement approach is based only on a sound processing method that produces a continuous shift in the HRTFs corresponding to the direction of the head, leading to a more advantageous way of simulating head rotations than the natural, real-world technique. In addition to this, the supplementary dynamic cues introduced in [13] are able to enhance localization, without requiring the listener to have control over the displacement of sound. The swing sound image method was initially developed by Kudo et al. [14], in order to improve spatial auditory perception in virtual environments that use digitally processed signals which were filtered with non-individualized HRTFs. Additionally, the sweep method is intended to reduce the incidence of front-back confusions through a continuous change in the values of the dynamic cues, such as spatial direction and time-based cues [13].

Early reflections effect on front-back disambiguation

It is believed that by adding early reflection to the sound played in a virtual auditory display, the listener will be given the perception of being immersed in a real-world acoustic environment (that contains reflections and reverberations) and thus improve his audio localization skills.

Perceptual feedback based training

In order to reduce the incidence of sound localization misjudgments (both angular precision errors and reversal errors), it has been demonstrated that a short period of acoustic adaptation to the altered hearing conditions is a necessary prerequisite that can significantly improve the spatial acoustic resolution in virtual auditory environments. Many sound localization experiment propose extensive training sessions, aiming at familiarizing the subject with the perception of 3D sounds (synthesized from non-individualized HRTFs) through a multimodal sensory experience [15] [16] [17]. The main advantage of this method is that the listeners learn how to adapt to the perception of generic sets of HRTFs, instead of modifying the spectral characteristic of the HRTFs to suit the auditory features of each listener apart [18]. A detailed description of the studies and experiments that focus on training the sound localization accuracy in virtual auditory environments through perceptual feedback based training can be found in [19].

The smallest head displacement that needs to be performed in order to record a significant reduction in the front-back confusion rate is of 2 degrees, when listening to pink noise sound stimuli, under free-field listening conditions [4]. At the same time, for broadband white noise presentation, a 4 degrees head rotation is sufficient for front-back disambiguation, although the highest accuracy is achieved when performing a displacement of 32 degrees [6]. However, for narrowband white noise stimuli, a movement of 10-20 degrees is mandatory [20].

In order to reduce the reversal errors, for both head and source movement experiments, under both virtual and free-field listening conditions, the subject is required to be aware and in control of the head, respectively source image spatial displacement [3]. For the source movement tests, the best swing angles to ensure efficient sound localization range between 4 to 8 degrees [14] [3] [21]. Nonetheless, in the experiment presented in [21], an angular displacement higher than 4 degrees is perceived as a moving source [3].

The swing sound image method (with the most effective swing angle of 8 degrees) reduces the incidence of front-back confusions (even for the non-individualized HRTF condition), as the swing presents a discrete, repetitive rendition of sound which makes the listener to become more aware of the direction of sound [14]. Correspondingly, the incidence of front-back confusions was lower for the moving auditory scene in both anechoic and reverberant environments [22]. Besides the swing method, the sweep approach reduced front-back confusions by 35% and 14% in the back [13].

In what concerns training as a method for improving localization judgment, white noise has proven to be a more efficient stimulus than pink noise. As a result, in the experiment described in [23], the front-back confusion rate reduced by 25% and the up-down confusion level decreased by 11% when listening to white noise, while when listening to pink noise, the rate of reversal errors decreased by 9.3% between the pre-test and the post-test sessions of the experiment [18].

The white noise was also effective for reducing front-back confusions in reverberant acoustic environments. The front-back confusion rate was lower in the reverberant condition than in the anechoic condition. This supports the belief that localization accuracy can be improved by adding reflections to a 3D auditory environment [22]. In consequence, azimuth localization errors are reduced in the full auralization condition. Thus, in the experiment presented in [22], the reversal errors percentage was lower in the reverberant condition than in the anechoic condition. By contrast, adding early reflections and reverberations to speech stimuli did not improve the front-back confusion rate.

Also, for other type of stimuli [24] [25], head movement, in combination with a reliable training procedure, succeeded to reduce the incidence of reversal localization misjudgments in virtual auditory environments between the pre-test and the post-test sessions of the experiments, from 59% to 28% [25] and from 11.2% (during the training procedure) to 8.5% (in the post-test phase) [24].

When performing a comparative study between the head movement and source movement methods, we can easily notice that a head displacement of at least 2 degrees (when listening to pink noise) and 4 degrees (when listening to white noise) is sufficient for reducing the front-back confusion rate. These results are close to those obtained in the experiments based on source image virtual movements (between 4 and 8 degrees), demonstrating that source image displacement can successfully replace head tracking methods in virtual auditory environments.

Moreover, it should be emphasized that sound source motions did not influence only the front-back perception, but they also conducted to a decrease in the rate of “in the head localization” errors [21]. Referring to the rate of sound externalization errors, it decreases in the reverberant condition (from 60% errors in the anechoic to 21% errors in the reverberant condition) [25].

DISCUSSION

Results of the experiments presented above conclude almost unanimously that head and body movements improve localization for the sound sources which are positioned in the front and rear hemifields. The rate of front-back confusions with virtual sources is higher than with real sources [3]. However, natural, self-initiated head rotations reduce the incidence of front-back confusions, even for virtual sources.

CONCLUSIONS

This paper presented a review of the most important methods designed to reduce the incidence of front-back confusions in both the free-field and virtual auditory environments. Our study concluded that head movements and sound source image movements are the most reliable techniques aimed at disambiguating sound localization in the front and rear hemifields. In addition to this, sound localization in reverberant environments offers the listener a complete perception of the extrapersonal space, including accurate judgment of the direction of the incoming sound in the case of audio sources positioned on the cone of confusion.

ACKNOWLEDGMENTS

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 643636 “Sound of Vision. This work has been funded by University Politehnica of Bucharest, through the “Excellence Research Grants” Program, UPB – GEX 2017. Identifier: UPB- GEX2017, 18 from 8.09.2017/2017 SAFE-VR.

REFERENCES

1. Ahveninen, J., Kopčo, N., Jääskeläinen, I.P., Psychophysics and Neuronal Bases of Sound Localization in Humans, Hear Res, 307, 86-97, (2014)
2. Zotkin, D.N., Hwang, J., Duraiswami, R., Davis, L.S., HRTF Personalization Using Anthropometric Measurements, Applications of Signal Processing to Audio and Acoustics, 157-160, (2003)
3. Wightman, F.L., Resolution of Front-Back Ambiguity in Spatial Hearing by Listener and Source Movement, J Acoust Soc Am, 105(5), 2841-53, (1999).
4. Hill, P.A., Nelson, P.A., Kirkeby, O., Resolution of Front-Back Confusion in Virtual Acoustic Imaging Systems, Journal of the Acoustical Society of America, 108(6), (2000).
5. Shinn-Cunningham, B.G., Santarelli, S., Kopco, N., Tori of Confusion: Binaural Localization Cues for Sources Within Reach of a Listener, J Acoust Soc Am, 107(3), 1627-36, (2000).
6. McAnally, K.I., Martin, R.L., Sound Localization with Head Movement: Implications for 3D Audio Displays, Frontiers in Neuroscience, 8, (2014).
7. Wenzel, E. M., Arruda, M., Kistler, D. J., Wightman, F. L., Localization using non-individualized head-related

transfer functions, Journal of the Acoustical Society of America, 94, 111-123, (1993).

8.Wightman, F.L., Kistler, D.J., Headphone simulation of free-field listening. I: Stimulus synthesis, J Acoust Soc Am, 85(2), 858-867, (1989).

9.Wightman, F.L., Kistler, D.J., Headphone simulation of free-field listening. II: Psychophysical validation, J Acoust Soc Am, 85(2), 868-878, (1989).

10.Reed, D.K., Maher, R.C., An Investigation of Early Reflection's Effect on Front-Back Localization in Spatial Audio, Proc. AES 127th Convention, New York, NY, USA, (2009).

11.Wallach, H., The Role of Head Movements and Vestibular and Visual Cues in Sound Localization, Journal of Experimental Psychology, 27(4), 339-368, (1940).

12.Mariette, N., Mitigation of Binaural Front-Back Confusions by Body Motion in Audio Augmented Reality, Proceedings of the 13th International Conference on Auditory Display, Montreal, Canada, (2007).

13.Towers, J., Burgess-Limerick, R., Riek, S., Improving 3-D Audio Localisation through the Provision of Supplementary Spatial Audio Cues, The Open Ergonomics Journal, Vol. 5, 1-9, (2013).

14.Kudo, A., Higuchi, H., Hokari, H., Shimada, S., Improved Method for Accurate Sound Localization, Acoustical Science and Technology, 27(3), 134-146, (2006).

15.Ahveninen, J., Jääskeläinen, I.P., Raij, T., Bonmassar, G., Devore, S., Hämäläinen, M., Levänen, S., Lin, F.H., Sams, M., Shinn-Cunningham, B.G., Witzel, T., Belliveau, J.W., Task-modulated "what" and "where" pathways in human auditory cortex, Proc. Natl. Acad. Sci. U.S.A. Proc Natl Acad Sci U S A, (2006).

16.Mendonça, C., A review on auditory space adaptations to altered head-related cues, Front Neurosci, (2014).

17.Fritz, J.B., Elhilali, M., David, S.V., Shamma, S.A., Does Attention Play a Role in Dynamic Receptive Field Adaptation to Changing Acoustic Silence in A1?, Hear Res., 229(1-2), 186-203, (2007).

18.Blum A., Katz B. F. G., Warusfel O., Eliciting adaptation to non-individual HRTF spectral cues with multi-modal training, Proceedings of Joint Meeting of the German and the French Acoustical Societies (CFA/DAGA '04), (2004)

19.Bălan, O., Moldoveanu, F., Moldoveanu, A., Negoi, I., The Role of Perceptual Feedback Training on Sound Localization Accuracy in Audio Experiments, Proceedings of the eLearning and Software for Education International Conference, 23-24 April 2015, Bucharest, Romania, (2015)

20.Macpherson, E.A., Cue Weighting and Vestibular Mediation of Temporal Dynamics in Sound Localization via Head Rotations, Proceedings of Meeting on Acoustics, Montreal, Canada, (2013).

21.Wersenyi, G.., Effect of Emulated Head-Tracking for Reducing Localization Errors in Virtual Audio Simulation, IEEE Transactions on Audio, Speech, and Language Processing, 17(2), 247-252, (2009).

22.Nykanen, A., Zedigh, A., Mohlin, P., Effects on Localization Performance from Moving the Sources in Binaural Reproductions, Proceedings of the 42nd International Congress and Exposition on Noise Control Engineering (INTER-NOISE 2013), Innsbruck, (2013).

23.Parseihian, G., Katz, B.F.G., Rapid Head-Related Transfer Function Adaptation Using a Virtual Auditory Environment, J. Acoust. Soc. Am., 131(4), (2012).

24.Roginska, A., Wakefield, G.H., McMullen, K., Searching for Sources from a Fixed Point in a Virtual Auditory Environment, Proceedings of the 17th International Conference in Auditory Display, Budapest, Hungary, (2011).

25.Begault, D.R., Wenzel, E.M., Anderson, M.R., Direct Comparison of the Impact of Head Tracking, Reverberation, and Individualized Head-Related Transfer Functions on the Spatial Perception of a Virtual Speech Source, J Audio Eng Soc., 49(10), 904-16, (2001).