Proiect cofinanţat din Fondul Social European prin Programul Operaţional Sectorial Dezvoltarea Resurselor Umane 2007-2013
Investeşte în oameni!
Proiect InnoRESEARCH - POSDRU/159/1.5/S/132395
Burse doctorale și postdoctorale în sprijinul inovării și competitivității în cercetare

UNIVERSITATEA **POLITEHNICA** DIN BUCUREŞTI
Facultatea de Automatică şi Calculatoare
Departamentul de Calculatoare

Nr. Decizie Senat 236 din 21.07.2015

TEZĂ DE DOCTORAT

Contribuţii la percepţia spaţială a sunetelor 3D şi la navigarea persoanelor cu deficienţe de vedere, prin antrenarea bazată pe feedback multimodal

Contributions to 3D sound-based space perception and navigation of visually impaired, through multimodal feedback training

Autor: Florina-Oana BĂLAN
Conducător de doctorat: Prof. Dr. Ing. Florica MOLDOVEANU

COMISIA DE DOCTORAT

Preşedinte	Prof. Dr. Ing. Adina FLOREA	de la	Universitatea POLITEHNICA din Bucureşti
Conducător de doctorat	Prof. Dr. Ing. Florica MOLDOVEANU	de la	Universitatea POLITEHNICA din Bucureşti
Referent	Prof. Dr. Ing. Florin RĂDULESCU	de la	Universitatea POLITEHNICA din Bucureşti
Referent	Prof. Dr. Ing. Pawel STRUMILLO	de la	Lodz University of Technology, Polonia
Referent	Prof. Dr. Ing. György WERSÉNYI	de la	Széchenyi István University, Györ, Ungaria

Bucureşti

ACKNOWLEDGMENTS

First and foremost, my deepest appreciation goes to Prof. Florica Moldoveanu, for supporting and encouraging me throughout the past three years. I am profoundly grateful for her kindness and patience, good advice and for the trust that she has invested in me. It has been an honor to meet her and to be her PhD student.

I would like to thank Prof. Alin Moldoveanu, for being the best mentor I could possibly have. I truly appreciate all his contribution of time, ideas and funding, to make my PhD experience productive and stimulating. His enthusiasm and hard-working nature have proven to be contagious and motivational and I am thankful for the wonderful example he has provided to me. He is undoubtedly my role model.

I will forever be thankful to Prof. Gabriela Moise and Monica Vlădoiu, for encouraging me to pursue the PhD studies. They gave me inspiration and motivation from the very beginning and supported me throughout the way. They are the persons who changed the trajectory of my career and enabled me to discover the satisfactions of the research efforts. My gratitude for them is everlasting and they will always occupy a very special place in my heart.

I would like to thank Prof. György Wersényi, for opening a door in Györ when I needed most – for warmly welcoming me at his university, helping me with the sound localization experiments and presenting the wonderful family of the International Community for Auditory Display.

Many thanks go to Hunor Nagy, who accompanied me 6 or 7 hours per day, for two weeks, translating and discussing with the visually impaired subjects – for his courtesy, good will and patience.

To my parents, for their unfailing love and unconditional sacrifice.

The work has been funded by the Sectoral Operational Programme Human Resources Development 2007-2013 of the Ministry of European Funds through the Financial Agreement POSDRU/159/1.5/S/132395.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

1.1. MOTIVATION

A significant problem that affects the lives of the blind and visually impaired people is sedentariness, the lack of working activities and of social involvement. An assistive device, based on an alternate sensory modality, such as hearing or touch, can help the blind individuals to gain spatial awareness, to enhance their navigational skills and to improve their quality of life. Nonetheless, in order to develop a reliable audio-based assistive device, we have to consider the users' sound localization performance and spatial auditory resolution.

Until now, many electronic devices that ensure assistance to visually-impaired people have been developed, but actually few of them are used. This is due to their lack of portability and wearability and to the limited quality and resolution of the detection, recognition and sensory-substitution components. The reluctance of the blind people to try to devices and the lack of organized training procedures is a significant obstacle that prevents the success of the electronic travel aids in the visually impaired community. The aim of sensory substitution systems is to restore the blind people the capacity to categorize and localize objects rapidly and to navigate in unknown environments. Such a device can address one of the major problems faced by the visually impaired individuals - assistance for their daily life, cardinality, autonomy, mobility and prevention of dangerous situations.

It has been demonstrated that the blind people possess adaptive skills associated with the remaining functional senses, such as hearing and touch. The neuroplastic changes occur in the anatomical regions of the brain that are related with the remaining senses and also in the occipital visual cortex. These findings do not exclude the possibility that the brain creates a visual and motor imaginary representation of the space, based on auditory and haptic cues.

Training plays an important role in improving the sound localization abilities of the blind listeners. The subjects who received intensive auditory training were able to better process the information and assess the visual principles of navigation and orientation. For example, an effective practice should concentrate on the employment of dynamic and interactive training procedures and on the use of functional tools such as virtual auditory environments or audio games.

As the virtual auditory displays and assistive devices employ 3D binaural sounds synthesized from non-individualized Head Related Transfer Functions (HRTF) which offer an ambiguous spatial auditory perception, we identified the need of training the visually impaired subjects' sound localization abilities through a perceptual feedback based learning approach. This research is oriented towards the study of binaural sounds, audio games and virtual auditory displays as rehabilitative tools for the visually impaired people. In particular, it is focused on developing effective methods for training the sound localization accuracy and spatial auditory perception of the blind people, based on perceptual feedback training and crossmodal sensory interaction.

1.2. GOALS OF THE RESEARCH

Worldwide, more than 285 million people suffer from a certain degree of visual impairment, of which 40 million are legally blind [1]. As a result, they need an assistive device that would help them to navigate in unfamiliar settings by improving their orientation and mobility skills. This study is part of a more complex research project which has as purpose the development of an assistive system for the visually impaired people that would replace sight with an alternative sensory modality, such as hearing and touch. The final prototype will employ 3D binaural sounds, vibrations and other auditory or haptic cues. The 3D binaural sounds incorporate directional information that defines the position of the sound source both in the horizontal and in the vertical planes, offering the same auditory perception as under free-field listening conditions. The 3D sounds are synthesized using the Head Related Transfer Functions, a measure of the sound transformation from the source to the listener's ears [2]. However, the HRTFs are highly dependent on the anatomical characteristics of the listener's body (size and shape of the pinna, head and torso), so that the use of the same HRTFs for different listeners will conduct to an ambiguous auditory perception and high localization errors. Due to the demanding and laborious process of recording individualized transfer functions for each listener apart, the majority of virtual auditory displays employ 3D sounds generated from non-individualized HRTFs that lead to a less accurate sound localization performance which is reflected in a higher incidence of precision and reversal errors [3], [4], [5].

This research is oriented towards improving the blind people's spatial auditory resolution through perceptual feedback based training, crossmodal sensorial adaptation (visual, auditory and haptic) and procedural learning. For this, we performed some experiments with the sighted and visually impaired people, in order to study and compare the level of spatial auditory performance they can achieve as a result of training. The theoretical and practical results obtained in this research will be used for creating an effective strategy that will improve the sound localization accuracy of the visually impaired people (in respect with the development of a sensory substitution device aimed at providing a rich representation of the environment) and for designing audio games that are targeted for both the blind community and for the sighted players.

1.3. THE AUTHOR'S SCIENTIFIC PUBLICATIONS IN CONNECTION WITH THIS THESIS

PAPERS PUBLISHED IN CONFERENCES

1. *O. Bălan*, *A. Moldoveanu, F. Moldoveanu, A. Morar*, "Realitatea Virtuală în Medicină - Realizări, Probleme şi Tendinţe", ROCHI National Conference on Computer-Human Interaction, Cluj-Napoca, September 2-3, 2013

2. *A. Morar, A. Moldoveanu, F. Moldoveanu, V. Asavei, L. Petrescu, O. Bălan, I. Negoi, S. Hostiuc*, "Interfețe Om-Mașină Bazate pe Realitate Augmentată/virtuală și Prelucrare de Imagini în Chirurgia Minim Invazivă", ROCHI National Conference on Computer-Human Interaction, Cluj-Napoca, September 2-3, 2013

3. *O. Bălan, A. Moldoveanu, F. Moldoveanu, A. Morar*, "Experiments on Training the Human Localization Abilities", Proceedings of the 10th International Scientific Conference eLearning and Software for Education-Bucharest, April 24-25, 2014, Vol. 2, ISSN: 2066 - 026X print 2066 - 8821 online **(ISI)**

4. *A. Moldoveanu, O. Bălan*, "Training System for Improving Spatial Sound Localization", Proceedings of the 10th International Scientific Conference eLearning and software for Education-Bucharest, April 24-25, 2014, Vol. 1, ISSN: 2066 - 026X print 2066 - 8821 online **(ISI)**

5. *O. Bălan, A. Moldoveanu, F. Moldoveanu*, "Studiu Comparativ al Principalelor Aplicaţii de Tipul Virtual Dressing Room", ROCHI National Conference on Computer-Human Interaction, Constanta, September 4-5, 2014, pp. 29-34, ISSN 2344-1690

6. *O. Bălan, A. Moldoveanu, F. Moldoveanu, M. I. Dascălu*, "Navigational 3D Audio-Based Game - Training Towards Rich Auditory Spatial Representation of the Environment", Proceedings of the 18th International Conference on System Theory, Control and Computing, Sinaia, Romania, October 17-19, 2014, pp. 688-693, ISBN 978-1-4799-4602-0 **(ISI, IEEE)**

7. *O. Bălan, A. Moldoveanu, F. Moldoveanu, M. I. Dascălu*, "Audio Games - A Novel Approach Towards Effective Learning in the Case of Visually-Impaired People", Proceedings of the ICERI Conference, Seville, Spain, November 17-19, 2014, pp. 6542-6548, ISBN 978-84-617-2484-0 **(ISI)**

8. *M. I. Dascălu, A. Moldoveanu, G. Dragoi, O. Bălan*, "Understanding and Improving the Usage and Impact of E-Learning in Medical Education", Proceedings of the ICERI Conference, Seville, Spain, November 17-19, 2014, pp. 6574-6580, ISBN 978-84-617-2484-0 **(ISI)**

9. *O. Bălan, A. Moldoveanu, A. Butean, F. Moldoveanu, I. Negoi*, "Comparative Research on Sound Localization Accuracy in the Free-Field and Virtual Auditory Displays", Proceedings of The 11th International Scientific Conference eLearning and software for Education, Bucharest, April 23-24, 2015. **(ISI)**

10. *O. Bălan, A. Moldoveanu, F. Moldoveanu, I. Negoi*, "The Role of Perceptual Feedback Training on Sound Localization Accuracy in Audio Experiments", Proceedings of The 11th International Scientific Conference eLearning and software for Education, Bucharest, April 23-24, 2015 **(ISI)**

11. *O. Bălan, A. Moldoveanu, F. Moldoveanu*, "Binaural Sound Analysis and Spatial Localization for the Visually Impaired People", 9th International Conference on Interfaces and Human Computer Interaction, 22 – 24 July 2015, Las Palmas de Gran Canaria, Spain **(ISI)**

12. *O. Bălan, A. Butean, A. Moldoveanu, F. Moldoveanu*, "Auditory and Haptic Spatial Cognitive Representation in the Case of the Visually Impaired People", The 22nd International Congress on Sound and Vibration, Florence, Italy, 12-16 July 2015

13. *A. Butean, O. Bălan, A. Moldoveanu, F. Moldoveanu*, "Touchscreen Based Audio and Vibro-Tactile Applications as Assistive Systems for People Suffering from Eye Disorders", The 22nd International Congress on Sound and Vibration, Florence, Italy, 12-16 July 2015

14. *O. Bălan, A. Moldoveanu, H. Nagy, G. Wersényi, N. Botezatu, A. Stan, R. G. Lupu*, "Haptic-Auditory Perceptual Feedback Based Training for Improving the Spatial Acoustic Resolution of the Visually Impaired People", The 21st International Conference on Auditory Display (ICAD-2015), July 8-10, 2015, Graz, Austria

15. *O. Bălan, A. Moldoveanu, F. Moldoveanu*, "3D Audio and Haptic Interfaces for Training the Spatial Acoustic Resolution in Virtual Auditory Environments", The 21st International Conference on Auditory Display (ICAD-2015), July 8-10, 2015, Graz, Austria

16. *O. Bălan, A. Moldoveanu, F. Moldoveanu, A. Butean*, "Developing a Navigational 3D Audio Game with Hierarchical Levels of Difficulty for the Visually Impaired Players", a 12-a Conferinţă Naţională de Interacţiune Om-Calculator, 24-25 September 2015, Bucharest, Romania **(ISI)**

17. *B. Troancă, A. Butean, A. Moldoveanu, O. Bălan*, "Introducing Basic Geometric Shapes to Visually Impaired People Using a Mobile App", a 12-a Conferinţă Naţională de Interacţiune Om-Calculator, 24-25 September 2015, Bucharest, Romania **(ISI)**

18. *A. Butean, O. Bălan, A. Moldoveanu, F. Moldoveanu*, "ICT Evolutions Supporting the Development of Assistive Systems for Visually Impaired People", WPA 2015 Bucharest International Congress 24 - 27 June, Bucharest

PAPERS PUBLISHED OR SUBMITTED TO JOURNALS

19. *O. Bălan, A. Moldoveanu, F. Moldoveanu, A. Morar*, "Assistive IT for Visually Impaired People", Journal of Information Systems & Operations Management, **vol**.7, no.2, 2013, pp. 391-404, ISSN 1843-4711

20. *O. Bălan, A. Moldoveanu, F. Moldoveanu, O.M. Ferche*, "O Scurtă Clasificare a Jocurilor Audio în Contextul Îmbunătăţirii Interacţiunii şi Accesibilităţii pentru Persoanele Nevăzătoare", Revista Română de Interacţiune Om-Calculator, **vol.** 7, no. 3, 2014, pp. 225-238, ISSN 1843-4460

21. *O. Bălan, A. Moldoveanu, F. Moldoveanu*, "Navigational Audio Games: An Effective Approach Towards Improving Spatial Contextual Learning for Blind People", The International Journal on Disability and Human Development, **vol**. 14, no. 2, April 2015, pp. 109-118, ISSN (Online) 2191-0367, ISSN (Print) 2191-1231, DOI: 10.1515/ijdhd-2014-0018

22. *O. Bălan, A. Moldoveanu, F. Moldoveanu*, "Spatial Auditory Representation in the Case of the Visually Impaired People", Journal of Information Systems & Operations Management, **vol.** 9, no. 1, May 2015, pp. 1-11.

23. *A. Butean, B. Troancă, O. Bălan, F. Moldoveanu, A. Moldoveanu, D. Chiriţă*, "Applications on Touchscreen Mobile Devices for Visually Impaired People", Romanian Journal of Human-Computer Interaction, **vol.** 8, no. 2, 2015, pp. 121-138

24. *O. Bălan, A. Moldoveanu, F. Moldoveanu*, "Spatial Sound Based System for Improving Orientation and Mobility Skills in the Absence of Sight", UPB Scientific Bulletin **(accepted for publication)**

25. *O. Bălan, A. Moldoveanu, F. Moldoveanu*, "Perceptual Feedback Training for Improving Spatial Acuity and Resolving Front-Back Confusion Errors in Virtual Auditory Environments", Archives of Acoustics **(in review) (ISI, Impact Factor = 0.65)**

26. *O. Bălan, A. Moldoveanu, F. Moldoveanu*, "A Systematic Review of the Methods and Experiments Aimed to Reduce Front-Back Confusions in the Free-Field and Virtual Auditory Environments", International Journal of Acoustics and Vibration **(in review) (ISI, Impact Factor = 0.37)**

27. *O. Bălan, A. Moldoveanu, F. Moldoveanu*, "Comparative Analysis on How Multimodal Perceptual Feedback Training Improves the Spatial Auditory Performance of the Sighted and Visually Impaired People", Archives of Acoustics **(in review) (ISI, Impact Factor = 0.65)**

28. *Ó.I. Jóhannesson, O. Bălan, R. Unnthorsson, A. Moldoveanu, Á. Kristjánsson*, "The Sound of Vision: On the Feasibility of Creating a Representation of the Auditory Environment for Spatial Navigation for the Visually Impaired", Frontiers in Human Neuroscience **(in review) (ISI, Impact Factor= 2.9)**

29. *O. Bălan, A. Moldoveanu, F. Moldoveanu, G. Wersényi*, Improving the Audio Game Playing Performances of the Visually Impaired People through Multimodal Training, Journal of Visual Impairment & Blindness (in review) **(ISI, Impact Factor=0.72).**

BOOKS

30. *O. Bălan, A. Moldoveanu, F. Moldoveanu*, ICT for Visually Impaired, Printech, ISBN 978-606-23-0348-8, 2014

PARTICIPATION IN DOCTORAL CONSORTIUM AT INTERNATIONAL CONFERENCES

- **Nordi'CHI** 2014, 26-30 October 2014, Helsinki, Finland (acceptance rate: 28%)
- **International Conference on Auditory Display (ICAD)**, 6-10 July 2015, Graz, Austria
- The 9th International Conference on **Interfaces and Human Computer Interaction**, 22 – 24 July 2015, Las Palmas de Gran Canaria, Spain (acceptance rate: 16%)

PRIZES AND AWARDS

- **Best Paper Award for PhD Students**, **The 18th International Conference on System Theory, Control and Computing (ICSTCC)**, Sinaia, Romania, 17-19 October 2014

RESEARCH VISITS

- **Széchenyi István University** from **Györ, Hungary** – 12.01.2015-23.02.2015 and 15.04.2015-18.05.2015. **Supervisor**: Prof. György Wersényi.

RESEARCH PROJECTS

- **Sound of Vision (Natural sense of vision through acoustics and haptics),** Horizon 2020, H2020-PHC-2014, project code 643636.

- **TRAVEE (Virtual Therapist with Augmented Feedback for Neuromotor Recovery),** PNCDII- Parteneriate, PN-II-PT-PCCA-2013-4-1580.

1.4. STRUCTURE OF THE THESIS

Chapter 2 presents some theoretical aspects of binaural hearing and the most relevant methods and experiments designed to decrease the rate of front-back confusion errors that appear when listening to 3D sounds delivered over headphones in virtual auditory displays or in the free-field. In addition, it contains a review of the auditory localization performance (angular precision localization error and front-back confusion rate) in both the free-field and in virtual auditory displays, together with the most important experimental approaches and learning procedures aimed at helping the listeners to adapt to altered hearing conditions (such as the use of non-individualized HRTFs for synthesizing 3D binaural sounds).

Chapter 3 performs a comprehensive parallel study on the most relevant sound localization experiments that involved the participation of both sighted and visually-impaired people (either suffering from congenital blindness or from late-onset visual impairments), with the purpose of understanding how the visual condition of a person influences his ability to perform spatial auditory localization tasks.

Chapter 4 contains a brief classification of the most significant audio games and discusses their functional aspects, means of interaction and design strategies. Also, it reviews the most notable navigational audio-based games in terms of their conceptual and technological approaches, accessibility and user interaction modality.

Chapter 5 details a comparative study on the degree of spatial auditory improvement achieved by a group of sighed and visually impaired individuals who participated in a sound localization experiment that used a multimodal perceptual feedback based training procedure. The sound localization performance was assessed before, after and during the training session. The results demonstrated that both the sighted and the visually impaired subjects succeeded in improving their spatial auditory resolution that has been reflected in a higher angular precision accuracy and a lower rate of front-back confusions.

Chapter 6 presents two experiments (one performed with sighted subjects and the other one with visually impaired individuals), conducted in order to investigate the improvements that occur in the sound localization performance and elementary environmental navigation skills of both groups of subjects as a result of training. The results of the experiments demonstrate a rapid adaptation of the subjects to the perception of 3D sounds synthesized with non-individualized HRTFs and enhanced sound localization and navigational skills as a consequence of training.

Chapter 7 summarizes the original contributions of the thesis and the future research perspectives.

CHAPTER 2
SOUND AND HEARING

Spatial hearing is an essential prerequisite that influences the human perception and behavior in complex communication environments [6]. Interacting with the perceptual elements of the extrapersonal 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. Moreover, auditory cues raise awareness about the surroundings, lead visual attention and transmit a wide range of sensorial information.

The sound is a very complex form of energy. 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. Perception is an automatic and unconscious process which transforms the auditory 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 auditory 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 delivered through headphones, the listener’s ability to locate the sound sources is lower than under real-world conditions. This is due to the problems that arise during binaural reproduction and simulation with non-individualized HRTFs, the lack of head dynamics (it is well known that 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 confusion rate.

The state of the art in the psychoacoustic field that has been presented in this chapter is included in [7], [8], [9] and [10].

2.1. ELEMENTS OF PSYCHOACOUSTICS AND HEARING PROCESSING

In terms of physics, the sound represents a series of vibrations that are transmitted through air or water, in the form of perceptible mechanical waves of fluctuating frequency, pressure and spatial displacement, which are capable of being recognized by the human hearing system. In psychology, sound is referred to as the acquisition and perception of audible information from the environment at the brain level. When we hear an audio signal, we commonly identify the amplitude (or loudness), frequency (or pitch) and the timbre of the sound.

Nonetheless, a prevailing question that the brain often generates is “where is the sound coming from?”. Knowing from where the sound source is originating is an important information. It enables the listener to become aware of his position in space and to discriminate from a multitude of incoming sounds, directs visual attention and assigns a source to a particular location. Audio localization cues are essential perception indicators in a large variety of real-world situations. Even if the visual system is 100 times more receptive to outer stimuli than the hearing system, the eyes are restricted to perceive only the information contained in our field of view, while hearing offers a complete 3D sensorial image of space [11].

This fact is supported by the values of the auditory Simple Reaction Time (SRT) parameter, which range from 100 to 160 ms in the case of the hearing sense and between 200 and 250 ms in the case of visual perception.

Spatial awareness represents the outright multisensory experience of being immersed into a real or virtual environment. It is a perceptual phenomenon which implies the subject to be fully conscious of his position, location and orientation in respect with the surroundings. The environment can be static or dynamic. In the latter case, the spatial perception is influenced by the displacement of the objects and events from the environment, the subject's movements or both. Spatial auditory awareness deals with the "presence, distribution and interaction of sounds with the surrounding objects" [12], audio detection, recognition, localization and discrimination. It is influenced by the listener's auditory capabilities, degree of familiarity with the environment or by the previous sensory experiences. According to [12], spatial auditory awareness has three major components: audio localization, distance estimation and spatial auditory assessment, which deals with the dimensions and features of the environment that affect audio distribution and human perception.

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 [5]. 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 and uniquely modify the frequency spectrum of each incoming sound. This specific auditory 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, the familiarity with the sound or the interference with the visual sense.

Hearing and auditory perception are impossible without the complex transformation of physical energy into active signals inside the ear. The auditory system performs a series of transmission, filtering and encoding tasks in order to build a complete perceptual image of the surrounding space. The sound entering the auditory system through the external and middle ear is processed and transformed into biological signals in the inner ear. The pinna (the external ear) and the concha (the hollow near the ear canal) gather and segregate the audio waves and transmit them to the middle ear through the external ear canal.

The perception of loudness (the auditory sensation correlated to the physical strength of the signal) [13] is produced by the variations in the intensity of sound. The human auditory system perceives loudness as a function of frequency. The highest sensitivity is acquired for frequencies exceeding 4 kHz. Moreover, loudness is also dependent on the duration of the stimulus. Short duration sounds (less than 100-200 ms) have a higher intensity than long duration stimuli. As a result, the exposure to prolonged sounds yields to a decrease of the perceived loudness and to an acute sense of dullness [14].

The perception of pitch is produced by the rate at which the audio waves vibrate at the eardrum. In fact, pitch is a subjective perceptual feature of sound that can be described as the impression by which the sounds can be ordered on a frequency-related scale, from low to high. Although pitch and frequency are often considered synonyms, there is a clear distinction between the two concepts, as pitch is not a physical property of the sound, but a highly individualized psychoacoustic feature of the perceived audio stimuli. There is a strong link between sound intensity and the perception of pitch. Thus, for low frequency sounds (less than 2 kHz), the level of pitch decreases with intensity, while for high frequency sounds (exceeding 4 kHz), the level of pitch is directly proportional to the audio intensity. In the latter case, the pitch increases as the sound gets louder [14]. There are two significant theories of pitch perception that describe how frequency – an objective, physical characteristic of sound is transformed in pitch – a subjective auditory perception by the complex mechanisms of the human auditory system. For instance, the place coding theory assumes that the pitch perception is acquired by the place of the highest stimulation on the basilar membrane. On the other hand, the temporal coding hypothesis states that the perception of pitch is influenced by the timing sound response model of the neural fibers that are connected to the cochlea [7], [15].

2.2. SPATIAL AUDITORY LOCALIZATION

2.2.1. Basic principles of spatial auditory localization

The main factors that influence spatial auditory perception are: horizontal localization (azimuth or the angle between the source and the median plane), vertical localization (elevation or the angle between the source and the plane that passes through both ears, incorporating the interaural axis), distance estimation and spaciousness (the perception of sound envelopment around the listener's body). Sound localization in the 3D environment ranges in task specificity from simple indications of the sound directions (front-rear, left-right) to relative discrimination (relative in respect to another sound source or point of reference) and absolute localization judgment (which consists of pointing exactly to the apparent direction of the incoming audio stimuli). In the category of relative localization measurements, we can distinguish the minimum audible angle (MAA), known also as "localization blur" [16], which represents the smallest audible difference in the position of two adjacent horizontal or vertical sound sources. For low-frequency broadband stimuli, the MAA is very small (around 1-2 degrees) in the front, raises to about 9-10 degrees laterally and presents a better discrimination accuracy (of 6-7 degrees) in the rear hemisphere. The largest value of the MAA is recorded in the mid-high frequency bands for lateral positions of over 40 degrees. In the case of absolute localization, the directional judgments are either made by indicating the perceived sound location (in degrees, all around the listener's head in the 3D space) or by selecting the apparent source from a list of predetermined directions (for instance, when the listener is required to identify the number of the active loudspeaker under free-field condition or when he is asked to specify the angular position using the clock time labeling – 12 for 0 degrees in the front, 3 for 90 degrees to the right etc.). The average absolute localization error is of about 5 degrees for the source positions situated in the front and 20 degrees for lateral angles [16].

2.2.2. Binaural cues

Binaural means listening with both ears, while monaural refers to perceiving auditory stimuli with a single ear [11].

The binaural cues define the angular direction of the sound source in the horizontal plane. As a result, they give information about where and how far from the median plane, on the left or right side of the listener's body, the incoming sound source is positioned. The binaural cues are unable to convey evidence about the elevation of the sound source or about whether it is located in the front, in the back, above or below the transversal (horizontal) plane [17].

If we imagine a sound source located on the right side of the head, it is easily noticeable to envision the path and the direction of the sound waves as they travel through air (with a velocity of 330 m per second). Thus, the sound waves will reach first the right ear (or the ear situated on the same side with the source - called ipsilateral ear) and then will diffract around the head to enter the left ear (called the contralateral ear). The arrival time depends on the length of the path from the source to both ears, including the distance that the sound needs to travel as it bends around the listener's head [11].

Low frequency sounds (below 800 Hz) are better localized by using the Interaural Level Difference (ILD), while higher frequency sounds (above 1.6 kHz) depend to a greater extent on the arrival time difference (Interaural Time Difference - ITD). They are the main binaural localization cues which guide the audio directional perception in the horizontal plane. However, there is an ambiguous transition area from 800 to 1600 Hz, where the ILD and ITD are observed simultaneously, offering poor localization accuracy [6]. The horizontal localization cues are the ILD and the ITD, as well as the spectral cues. ILD and ITD are associated to the directional judgment of sound localization, while the spectral cues relate to the front-back disambiguation [18] (Fig. 2.1.).

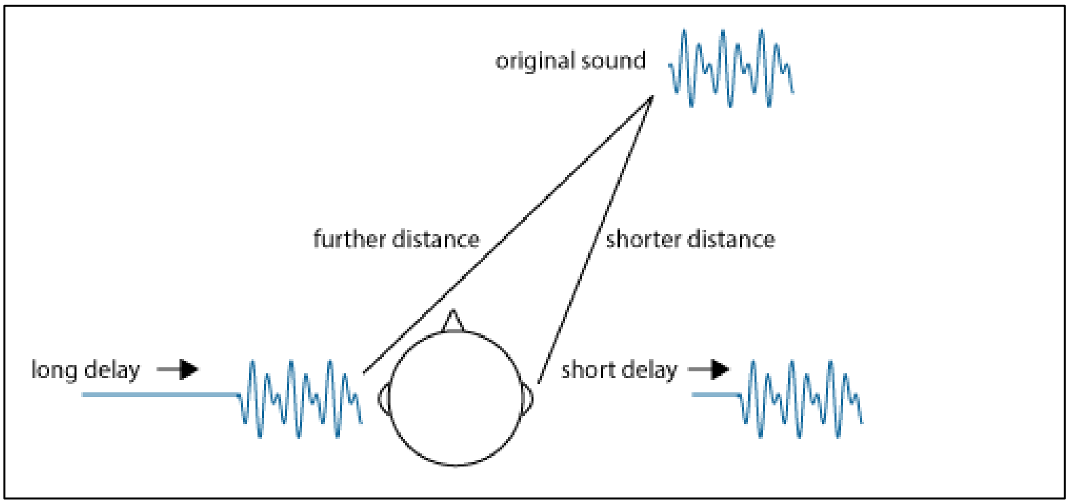

Fig. 2.1. The Interaural Time Difference [19]

THE INTERAURAL TIME DIFFERENCE

There are various formulas for calculating the ITD. Woodworth's formula (1954) presumes a frequency-independent design of a signal crossing a sphere-shaped head model:

$$ITD = \frac{r}{c}(\theta + sin\theta) \tag{2.1}$$

Where r is radius of the subject's head, θ is the angle between the listener and the sound direction and c is the speed of sound, of approximately 340 m/s at sea level. This formula can be applied only for high frequencies (above 3kHz), whereas for low frequency audio that presents longer waves tending to diffract around the listener's head the Kuhn formula (1977) is more applicable.

$$ITD = \frac{ar}{c} sin\theta \tag{2.2}$$

Where a=3 for frequencies below 500 Hz, a=2 for frequencies above 2 kHz and progressively decreasing for higher frequency levels. The ITD increases with a larger head size and decreases with higher sound frequency levels [11], [12], [14].

For any sound source located on the median plane, the ITD will be zero. The maximum ITD is obtained if the sound source is positioned at 90 degrees to the left or to the right. The highest value for the ITD is around 0.65 milliseconds, which represents the ratio of the distance between the ears to the speed of sound. The smallest detectable ITD that has been recorded is of 10 microseconds (for broadband noise in the horizontal plane), corresponding to an angular displacement of 1 degree (the minimum audible angle) from the forward or backward directions (at 0 degrees in the front and 180 degrees in the back) (Fig. 2.2.) [11], [12].

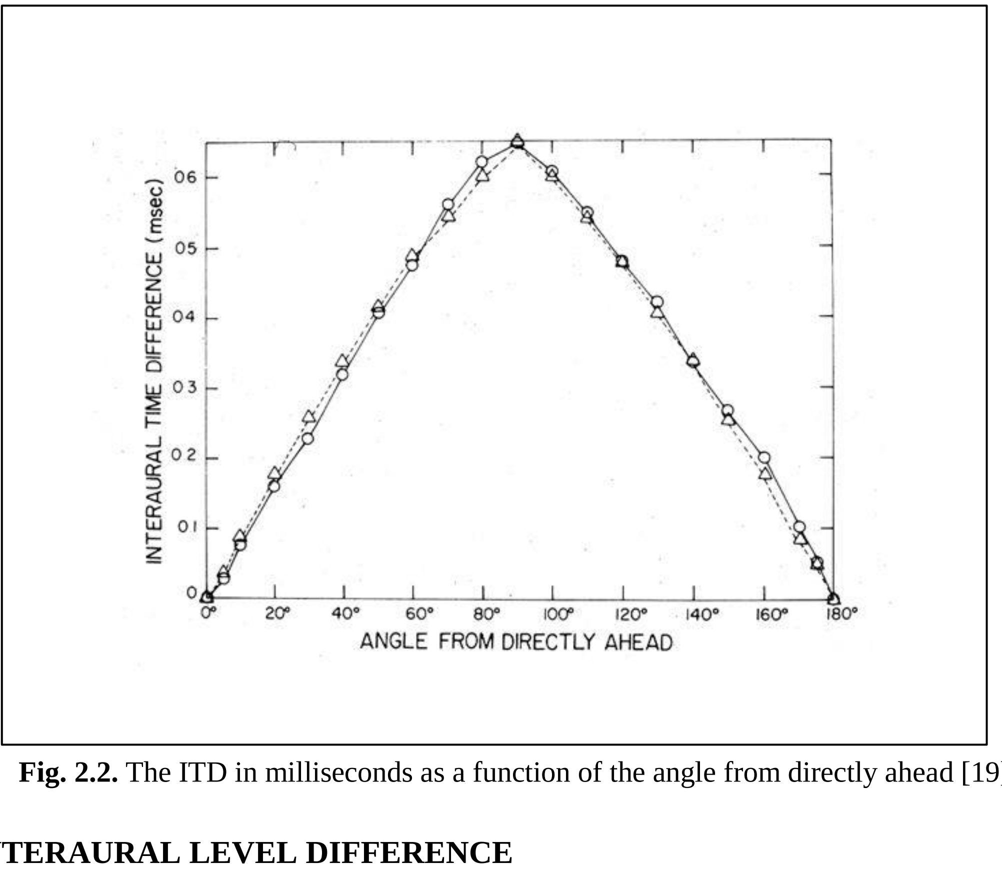

Fig. 2.2. The ITD in milliseconds as a function of the angle from directly ahead [19]

THE INTERAURAL LEVEL DIFFERENCE

The Interaural Level Difference (ILD) is caused by both the sound intensity variation with distance (the sound amplitude level increases as the sources get closer to the head) and by the "shadowing" effect of the head, which will cause diffractions of the waves, so that some amount of audio energy from a source located on the right side will be lost on the way to the left ear. Low frequencies (which have longer wavelengths) have the tendency to diffract more around the head, reducing the values of ILD (1 dB for a 200 Hz tone and 20 dB for a 6000 Hz sound). Thus, if in a complex, broadband sound, the level of high frequency components is higher, an increase in the values of the ILD will conduct to better localization accuracy. However, the minimum ILD detectable by the human hearing system lies in the range of 1-2 dB [11].

ILD depends on the lateral direction and frequency of the sound source. For instance, it has values ranging from 10 dB at 3 kHz frequency and of 35 dB at 10 kHz. ILD increases with frequency level and azimuth angle, reaching the maximum angle at 90 degrees in the horizontal plane.

Although the binaural cues are effective for sound localization and spatial discrimination in the horizontal plane, the left-right symmetry of the head causes a region of spatial ambiguity, called the “cone of confusion” (Fig. 2.3.). It can be described as an imaginary cone situated along the interaural axis, containing positions which present identical interaural differences. ITD and ILD are not constant for the sound sources located in the cone of confusion, mainly because the head is not a perfect sphere, the ears are not perfectly symmetrical and the interaural axis does not pass accurately through the middle of the head and ears [11], [12], [14], [21].

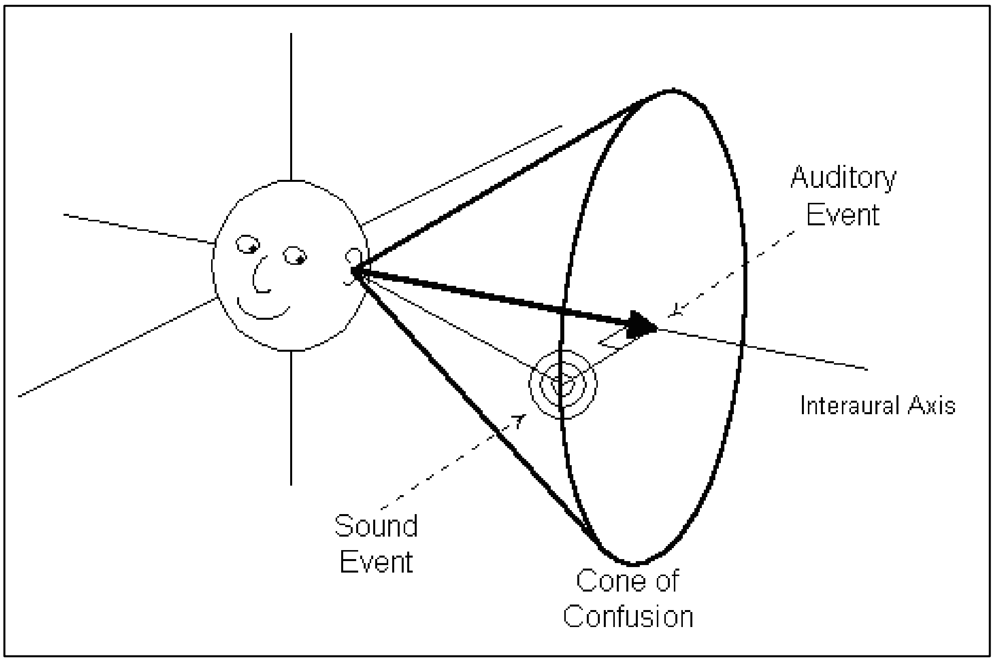

Fig. 2.3. The cone of confusion [20]

2.2.3. Monaural cues

In real-world conditions, the human auditory system makes use of other indicators, called monaural cues, which do not depend on the audio stimuli perception at the two ears. The monaural cues appear as a result of the energy absorption, shadowing effect and reflective properties of the head, torso, shoulders and pinna (the outer ear) that occur before the sound enters the ears.

The ridges, cavities and the other irregularities of the pinna produce modifications in the spectrum of the sound, such as reflection, absorption and resonance. As the cavities are relatively small, they only influence high-frequency sounds with short wavelengths, causing peaks and notches at frequency rates of over 4000 Hz (Fig. 2.4.).

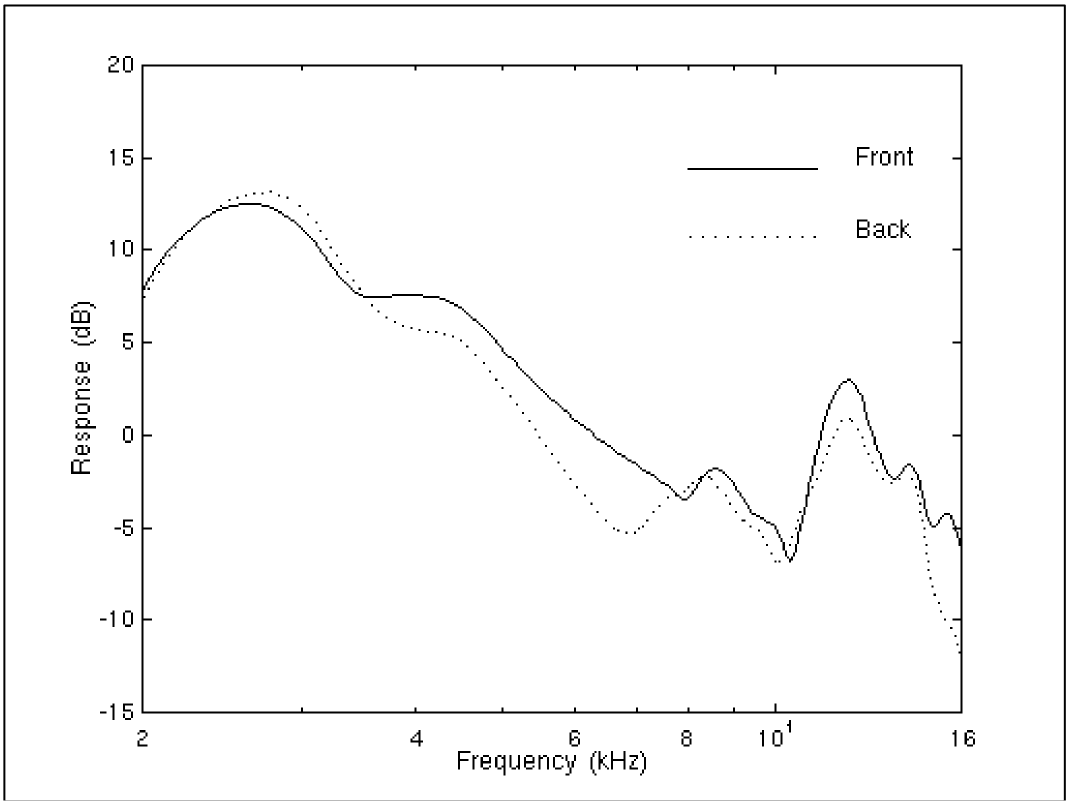

Fig. 2.4. Peak at 4 kHz (ear canal resonance) and pinna notch at 10 kHz that changes with elevation [22]

However, as the shape of these peaks and notches are dependent on the azimuth angle and direction of the sound source, they can be used as additional cues for localization in the horizontal plane. There is a shadowing effect of the pinna for the sound sources situated in the rear hemifield. In consequence, it produces a diffraction of the audio waves around the ear, especially for high frequency stimuli which tend to attenuate as a result of energy dissipation in the environment [11] (Fig. 2.5.).

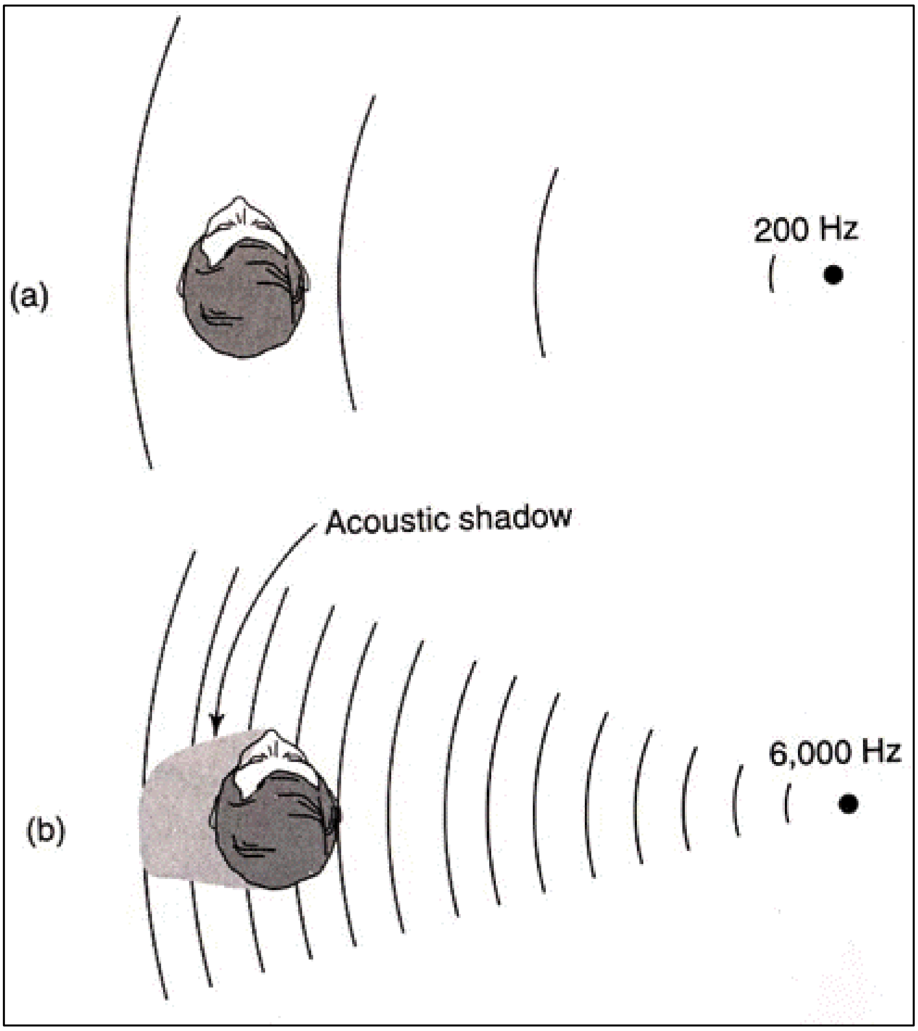

Fig. 2.5. The shadowing effect of the head [19]

The auditory shadow caused by the reflections of the sound behind the head is highly evident at frequencies of over 1 kHz. These effects cause modifications in the spectral content of the sound and produce peaks and notches at different frequencies and spatial directions. The monaural cues offer a better localization accuracy not only for the points situated inside the cone of confusion, but also for the sources positioned in the vertical plane (vertical localization), especially for high-frequency sounds. The pinna cues (which are more reliable for frequency ranges above 3-4 kHz) are essential also for the "out of the head" impression which conveys a spatial and in-depth distance perception of the emitting sound and also for localization in the horizontal plane, especially at frequencies of over 5 kHz. Nonetheless, vertical localization is, to some extent, less accurate than horizontal localization. The most relevant frequency ranges for front-back and up-down sound discrimination are between 4-16 and 6-12 kHz. For the median plane localization, there are 2 peaks (between 7-9 kHz and 10-12 kHz) and one notch (4-8 kHz), which are efficient for 3D audio playback in the virtual auditory environment.

Vertical localization is determined by the interaction of the sound with the pinna that produces a monaural cue with a particular spectral notch rising from 5 kHz to 10 kHz, as the sound is displaced from 0 degrees (at the level of the listener's head) to 90 degrees (above the head) [6], [7], [11].

THE HEAD RELATED TRANSFER FUNCTION

Sound scattering caused by the listener's body plays an important role in localization. As the sound waves travel through air, they encounter various obstacles until reaching both ears. In consequence, the audio signals are scattered, diffracted and reflected by the pinna, head, shoulders and torso, producing characteristic modifications in the spectrum of the sound [5].

The Head Related Transfer Function (HRTF) is a response which determines how the ear perceives the audio stimuli arriving from a particular point in space. It is a frequency-dependent transfer function describing the direction of the sound waves between the source and the ear (through the auditory canal, to the eardrums). The pair of HRTFs corresponding to the left and right ear designates how the sound source is being perceived by the listener [4] (Fig. 2.6.).

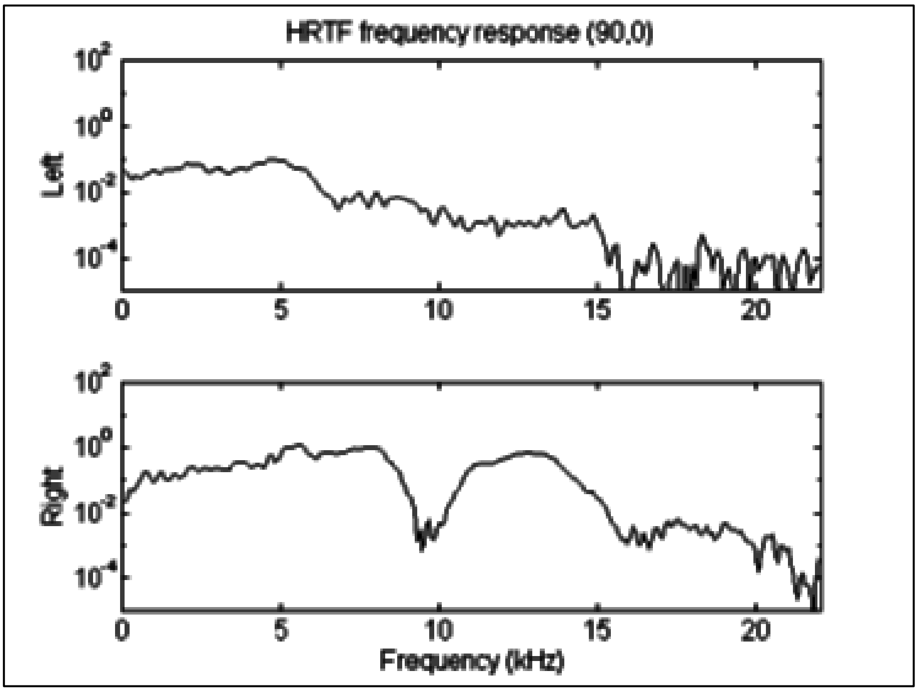

Fig. 2.6. Example HRTF magnitude for 90º azimuth, 0º elevation [23]

If the HRTF is known, it is easy to synthesize a scene of a virtual auditory environment that conveys to the listener the same perception as a real source in the real environment would produce [3]. In anechoic conditions, the filtering of the audio signal with the corresponding pair (left and right) of HRTFs that are characteristic to a direction in space delivers the same auditory pressure as listening to the same sound in a real-world environment. Moreover, the sensation of being immersed in the exocentric space is produced not only by filtering the sound with the HRTFs of the direction of the source, but also by introducing reverberations, reflections and movement cues that deliver a more realistic perception of the virtual environment [5].

The HRTFs are highly individualized, depending on the anatomical features, the head asymmetry and the ear placement of the listener. The specific differences in the anatomy of the ears, head and body (concerning both size and shape) do not allow for the use of the same HRTFs for all the listeners [3], [4]. As a result, the characteristics of the spectrum shape of the sound that point to certain source locations vary among the subjects, leading to ambiguous interpretations of the same HRTFs. Thus, the HRTF of one listener will conduct to an altered auditory perception when employed by other individuals. The theoretical approach of

Middlebrooks et al [5] argues that there can be a correspondence between the size of the pinna and the frequency spectrum of the HRTFs, so that the HRTFs can be calculated based on the variation of the outer ear dimensions among individuals.

Moreover, the HRTFs are not transferrable between individuals, so that they will not be perceived identically by all of the listeners, especially in the frequency region of 5-10 kHz, where the pinna effects are more prominent. Nonetheless, due to the difficulty in obtaining individualized HRTFs, the majority of sound localization experiments use generic, non-individualized HRTFs which are stored in large databases. However, generic HRTFs cause a high incidence of front-back confusions, perception distortion, degraded localization accuracy and a large number of spatial auditory discrimination errors.

The ideal virtual auditory display should employ individualized HRTFs for processing 3D sounds rendered over stereophonic headphones, because the sound perception at each ear can be controlled separately. Although, obtaining a full database of HRTF measurements is a time-consuming and unpractical task, as it requires recording the sounds that arrive at the eardrums by small microphones placed inside the ear of a human subject or of a dummy head mannequin from all the positions in the horizontal and in the vertical planes [2], [24] (Fig. 2.7.).

Fig. 2.7. HRTF recording [25]

HRTF PERSONALIZATION

HRTF customization is highly required in order to obtain improvements in the localization accuracy and individual perception in virtual spatial auditory environments. The HRTF personalization techniques are: the classical solution based on the measurement (in anechoic conditions) of the impulse response at both ears by using moving loudspeakers and small microphones placed inside the ears of the listener, numerical methods that calculate the scattering of the sound waves through the air and the anthropometrical approaches suggesting the existence of a level of interdependence between the pinna anatomical characteristics and the spectral shape of the HRTF [3], [4].

2.2.4. Dynamic cues

The binaural cues are efficient for sound localization in the horizontal plane, but unreliable for audio discrimination in the entire 3D space. For instance, the sound sources which are positioned in the median plane (in the front or in the rear hemifield) will produce a null ITD and ILD. As a result, it is impossible to predict whether the source is located in the front, back, above or below the head level. The dynamic cues are used to improve sound localization and to reduce the rate of front-back confusions (by changing the values of both ITD and ILD), in order to give the listener the necessary cues for identifying the exact direction of the sound source inside the ambiguous cone of confusion.

Besides the binaural and monaural cues, human sound localization and source discrimination are dependent on the head's dynamics, which causes modifications in the spectrum of the sound at each ear, particularly for frequencies below 2 kHz. Head movements improve localization and reduce errors and front-back confusions, especially for sound durations of 600-800 ms.

Thus, localization ambiguities are reduced by the modifications in the values of both the ITD and the ILD, which change as a result of the rotation of the listener's head relative to the position of the sound source. For instance, if a sound source is located in front of the listener and he makes a short movement to the right, the binaural values in the right ear will decrease, causing the sound to arrive firstly at the left ear. By contrary, the same head rotation will reduce the ITD and ILD rate in the left ear if the sound source is positioned in the back, producing a better detection at the right ear [11].

The increase in the signal's duration conducts not only to a more correct front-back discrimination, but also to a better localization accuracy. It is preferable for the incoming sound to be composed of repeated beats or short pieces of audio, rather than a single, brief signal. The listener's predisposition is to perform small back-and-forth head rotations in the horizontal plane, until reaching complete focus on the direction of the sound. However, quick head movements during the playback of a short audio stimulus conduct to localization errors and audio discrimination ambiguities. Wallach (1940) [17] recognized the remarkable role that head movements play for front-back disambiguation, due to the changes of the ITD and ILD during head displacement. It has been demonstrated that dynamic ITD changes are more efficient for reducing front-back confusions during head movement than the ILD [26], [27]. Moreover, he demonstrated later that head movements are not absolutely necessary for accurate front-back localization. Wallach moved the chair on which the listener was seated, without actually moving the head, suggesting that not the anatomical features of the head and torso are involved in the process of disambiguation, but the shift in the body's posture, direction and orientation [7].

2.2.5. Reversal localization errors

Externalization errors take place when the listener perceives the sound as coming from inside the head, lacking spatialization and distance perception. Similarly, it is called "inside-the-head localization", "lateralization" or "intracranial localization" [28]. This is due to the fact that the spectrum modifications related to the head and pinnae individualized features are not included in the audio signals that interact with the auditory system. It is important to perceive these transformations in order to get the sensation of externalization.

The study of Begault et al [28] showed a 7:1 decrease in front-back confusions and a 2:1 reduction in back-front confusions for head motion based localization in a virtual auditory environment rendered over broadband white noise audio stimuli. Moreover, reverberation and sound reflections are efficient means for resolving in-the-head localization by creating a natural, echoic environment resembling natural and real-world conditions.

CATEGORIES OF REVERSAL 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. For instance, for a sound source located at 30 degrees in the frontal hemisphere, the listener can perceive an apparent position at 150 degrees in the rear hemisphere [27].

Reversal errors are caused by the spherical shape of the head, by the reflections of the environment, by the interferences with other objects or sounds and by the modifications in the spectrum of the waves, caused by identical levels of ITD and ILD, especially under the headphones playback condition. 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" [27]. 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 [29]. Front-back confusions (Fig. 2.8.) 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 that are presented through headphones.

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 components of the environment by using the hearing sense as the main sensorial 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 is able to simulate the change in audio perception which accompanies the listener's head movements [29]. 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 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.

Headphones renderings of the sound in these types of applications lead 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 the sound localization cues are highly individualized and user-specific. To ensure accurate sound localization, measurements of the HRTF for each listener are 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 planes [26].

Sound synthesis using individualized HRTFs would be the best solution to simulate a free-field listening experience in a virtual auditory environment. Wightman [27] demonstrated in their experiments that the sound localization performance of 8 users in free-field listening conditions are comparable with the results obtained when the same listeners were presented 3D binaural sounds synthesized from their own HRTFs, under headphones playback conditions. Nonetheless, some small localization errors did appear in the second experiment (the rate of front-back confusions increased from 6% to 11%), suggesting that the headphone rendering approach introduces minor, but noticeable inaccuracies.

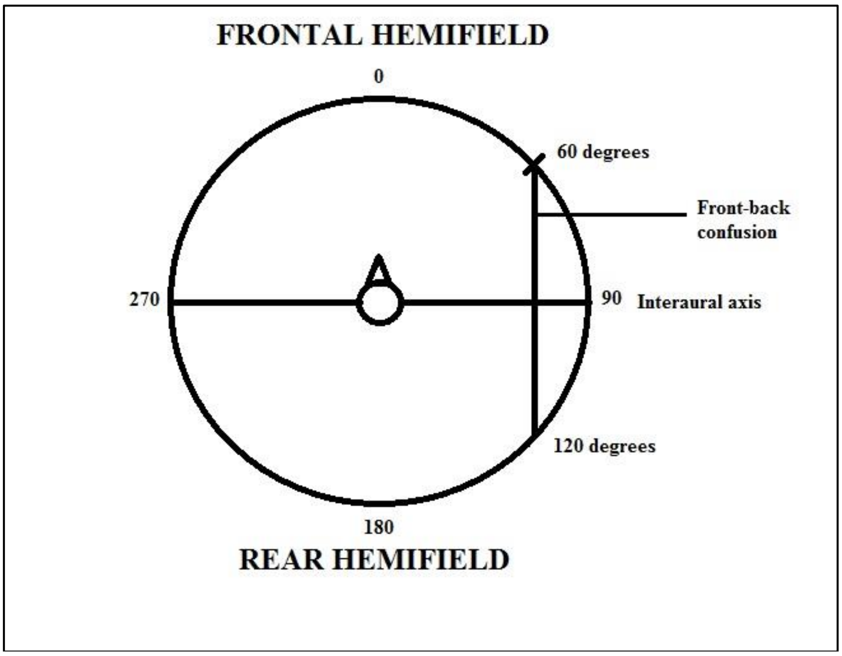

Fig. 2.8. Example of front-back confusion

2.2.6. Methods to reduce front-back confusions

HEAD MOVEMENT

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 [17] 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. During head movement, there are 3 categories of sensory information that affect localization perception: proprioceptive sensations obtained from the neck muscles involved in the process of movement, visual information and the stimulation of the vestibular apparatus [7].

BODY MOVEMENT

The study described in [30] 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 positions. 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 demonstrated 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 positions) of 12 or 16 degrees offered the best rate of front-back accuracy. As a result, the dynamic body cues are able to reduce the incidence of reversal errors and to improve localization on the cone of confusion [7].

SOURCE MOVEMENT

In the sound localization experiments in which head movements are allowed [18], [31], [32], 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. In addition to this, the supplementary dynamic cues introduced in [32] are able to enhance localization, without requiring the listener to have control over the sound displacement.

EARLY REFLECTIONS

It is believed that by adding early reflections to the sound played in a virtual auditory display, the listener will be given the perception of being immersed in a real-world auditory environment (that contains reflections and reverberations) and therefore improve his audio localization skills [7], [28], [33].

2.3. SPATIAL AUDITORY ACCURACY AND LOCALIZATION PERFORMANCE IN THE FREE-FIELD AND IN VIRTUAL AUDITORY DISPLAYS

The sound localization tests that evaluate the human spatial discriminatory abilities vary considerably in their methodological approaches, pursued objectives, experimental procedures and localization paradigms. In virtual auditory displays, the listeners are presented various types of sounds via headphones (clicks, noises, pure tones) and are required to indicate the perceived lateral position of the sound source relative to the forward orientation of the head. This type of experimental procedure is called “lateralization” (judgment of phantom sound sources located in the head [16]) and it has therefore a different approach than the term “localization”, which stands for the apparent perception in both direction and distance (depth perception and externalization into the listener’s surrounding space) [34].

The human auditory system is able to combine the binaural and monaural cues into a unitary, consistent auditory image of space. Nevertheless, it has been argued that this information is acquired with the help of the visual sense, which offers sensory feedback and regulates the way in which the brain builds up the entire spatial representation of the environment [8], [35].

2.3.1. Simulation in the free-field and in virtual auditory displays

The free-field simulation involves the presence of several loudspeakers (disposed in an array or following an arrangement rule) around the listener, in both the horizontal and the vertical planes. In a virtual auditory display, the sound offers a naturalistic and dynamic representation of the scene objects from all the possible directions in space, improving the sense of presence, immersion and realism. Virtual auditory displays are useful for a wide range of applications, such as air traffic control displays, teleconference environments, assistive devices for the visually impaired people and audio games.

A dynamic and interactive auditory environment can be simulated by employing a head-tracking device that enables the users to move their heads and sequentially hear a change in the perceived sound that corresponds to the direction of the head displacement [8].

2.3.2. Spatial auditory accuracy and localization performance

According to [16], localization errors consist of two types of localization uncertainties: random errors (defined as the rate of imprecision caused by variability in the listening conditions) and constant errors (inaccuracies determined by the rough estimations of the direction of sound by the human auditory system).

Localization accuracy relies on the health of the human auditory system, the symmetry of the head, the quality of the sound, the listening conditions (spaciousness, reverberation, reflections, noises) and on the functionality of the electronic audio rendering system. The localization bias depends also on the data acquisition approach and on the pointing methods used by the subjects (finger pointing, head pointing, verbal indication) which clearly introduce a certain degree of variability.

2.3.3. Methods used for sound localization

The main methods for indicating the perceived sound location are head and hand pointing. Head pointing measurements are possible due to the existence of head tracking devices that are accessible, precise and easy to use. According to [36], turning the face towards the sound source is a very ecological action, as it adds the source into the visual field. Moreover, an advantage of the head pointing technique lies in the fact that the spatial coordinate system of the pointing method is concentric with that of the source [37]. Some inherent errors can occur when the subjects move their eyes to locate the source and deviate from the direction pinpointed by the tip of the nose. For instance, when they are required to identify sources situated on the extreme sides of the visual field, they have a natural tendency to rotate their eyes instead of directing the head. Moreover, as the head tracking device is mounted on the head and not at the eyes level, the subjects seldom misjudge the amplitude of their movement and bring the source into focus using the eyes. This systematic bias of direction estimation can be resolved by designing a training session that would familiarize the subjects with the pointing procedure and the potential localization error that can occur during the tests. In the experiment presented in [37], the subjects got accustomed with the requirements of the pointing task and learned very rapidly to orient using their heads. As a result, the response error introduced by head movements was significantly lower than the localization uncertainty produced by the perceptual misjudgment of the direction of the sound source. The same experiment demonstrated that there is a certain degree of uncertainty for localization in the rear hemifield when the head pointing method is used, mainly due to the difficulty in performing large rotations around the neck and waist [8].

2.4. PERCEPTUAL FEEDBACK TRAINING FOR IMPROVING SOUND LOCALIZATION ACCURACY IN AUDIO EXPERIMENTS

2.4.1. The effects of training on improving sound localization accuracy

In order to reduce the incidence of localization errors, it has been demonstrated that a period of adaptation to the distorted auditory rendering conditions is required, so that listeners would get accustomed with the demands of the new hearing experience (different temporal and spectral characteristics from the ones expected by the listener’s auditory system, in-the-head localization or inaccurate distance perception). In addition to this, the majority of sound localization experiments take into account trial or training sessions (generically referred to as the “learning effect”), aimed to familiarize the subjects with the perception of 3D sounds (especially in the case of inexperienced subjects), the requirements of the localization task and the pointing technique. The particularity of this method lies in the fact that the listeners learn how to adapt to non-individualized HRTFs, instead of modifying the HRTFs to suit the auditory characteristics of each listener apart [9], [38].

2.4.2. Perceptual learning

According to Goldstone [39], [40], perceptual learning represents the performance obtained from practice or experience that improves a person’s capacity to interact with the environment. It has been demonstrated in different experiments that sound localization accuracy can be enhanced through perceptual learning, training and behavioral adaptation [39].

The brain has a remarkable capacity to adapt to new auditory conditions and to learn to perform localization and discrimination tasks under altered spatial acoustic circumstances.

The human auditory system is highly flexible in what concerns the execution of spatial auditory encodings. The spatial sensory representation is continuously modeled by experience-driven plasticity and by regular adaptation to various auditory contexts, such as altered audio cues or modifications in the spectral and temporal characteristics of the sound. This adaptive process is the result of the brain's response to the conditions imposed by real-world situations (which involve the presence of dynamic or attention-dependent stimuli), emphasizing the brain's ability to accommodate to altered auditory inputs caused by unilateral or bilateral hearing distortions [39].

2.4.3. Experiments on improving sound localization accuracy

Sound localization performance can be altered by using ear blocks (unilateral blocks, where one ear is plugged with a sound attenuating earplug), ear molds (where wearable molds are inserted into the subjects' ears in order to produce anatomical transformations to the shape of the pinnae and to induce spectro-temporal modifications to the sound perception), electronic hearing devices (such as hearing aids and cochlear implants that require a significant period of adaptation) or the use of 3D sounds synthesized with non-individualized HRTFs in virtual auditory environments [41].

As Mendonça identified [41], there are 3 main training methods aimed at enhancing the sound localization performance under altered listening conditions. The sound exposure paradigm implies that the subjects implicitly, spontaneously and unconsciously learn to adapt to altered auditory cues with multisensory feedback, without being aware of the adaptation process. Sound exposure can be divided into long-term (where the modifications in the auditory cues are constant throughout the whole duration of the experiment) and intermittent exposure, where the changes in the localization cues are applied only in the experimental sessions. The "training with feedback" paradigm is composed of a pre-test session (with the purpose of assessing the subjects' initial spatial localization performance), a training session (usually consisting in a sound localization task that provides perceptual feedback concerning the correct direction of the auditory stimuli in space) and a post-test session, similar or identical with the pre-test phase, in which the degree of adaptation-based spatial localization improvement is evaluated. Moreover, in the third training paradigm, the "active learning" category, the subjects are actively involved in the process of spatial auditory adaptation, such as in Parseihian and Katz's experiment [42] where the participants were required to explore a virtual auditory environment by playing a game in which they needed to identify the directions of various sound sources using a hand held tracked ball.

SPECTRAL ALTERATION THROUGH MORPHOLOGICAL EAR MODIFICATIONS

The simplest method for testing the human auditory system's degree of adaptation to degraded auditory cues is based on inserting ears plugs in one ear [43], [44], [45], [46], [47]. In Florentine's experiment [44], the subjects were required to wear the unilateral block for a time period ranging from 5 to 101 days, Bauer [43] inserted the ear block for 65 hours, van Wanrooij and van Opstal [48] produced a monaural spectral disruption for 9 to 49 days, while Held [49] applied an electronic hearing device that presented sounds with displaced azimuths of 20 degrees, for 8 hours daily, constantly monitoring the degree of spatial auditory adaptation in specifically-designed periodical tests. When the spatial auditory performance has been measured following the ear plug insertion, the level of sound localization that has been recorded was very low, but it recovered significantly over a period of several days (Bauer, et al. – 2, 3 days, Florentine – 4-10 days, Kumpik, et al. – 7 days and no improvements have been found for shorter periods of plugging – up to 24 h – Slattery & Middlebrooks) [50]. In Held's experiment [49], the subjects recorded a displacement of the auditory representation halfway in the direction of the sound source's azimuth shift [41]. Moreover, Slattery and Middlebrooks [46] applied a monaural spectral displacement of 30 degrees to the side of the open ear for a period of 24 hours. However, the results recorded by the two groups of subjects who participated in the tests (a groups of normal listening individuals and another of people suffering from unilateral deafness) do not reveal a significant improvement in the sound localization accuracy and spatial auditory recalibration, mainly due to the short exposure period and to the lack of any type of training.

In Hofman's experiment [35] four sighted sighted subjects were required to wear custom-made molds inside the concha of both ears for 6 weeks. After the application of the molds, the sound localization accuracy in the vertical plane has been severely impaired. However, after several days, auditory localization performance improved gradually for all the subjects. After the molds have been removed, sound localization reached the same levels of accuracy recorded before the start of the experiments. The authors consider that the subjects developed two distinct spatial maps: one for the normal and the other for the altered spectral cues, in a way that resembles the acquisition of foreign languages in multilingual individuals. Similar patterns of multi-mapping processing have also been identified in the midbrain spatial localization pathway of the barn owls [51], [52].

Carlile and Blackman [53] investigated the rate of adaptation to new spectral cues for 76 equally distributed sound sources located inside and outside the listener's field of vision, using small bilateral ear molds inserted into the outer ear [50]. Immediately after the molds' insertion, the rate of front-back confusions increased 7 times, while the localization errors in the vertical plane doubled. After the 40.5 day period in which the listeners have been required to continuously wear the molds, the rate of front-back confusions, the polar angle error (elevation error on the cone of confusion) and the lateral (azimuth) error decreased significantly. Consequently to the molds removal, the spatial resolution returned to the initial values recorded before the molds' insertion - in complete accordance to the results obtained by Hofman [35], demonstrating that even after a large period of exposure to altered listening conditions, the brain still preserves the representation of the "old" spectral cues [10].

PERCEPTUAL FEEDBACK-BASED TRAINING

In the case of the "training with feedback" paradigm, the response feedback is usually provided as a "correct" or "incorrect" notification or as an explicit indication of the accurate location of the incoming sound source. For instance, Butler trained the sound localization abilities of the subjects participating in his experiment in the horizontal plane and offered them response feedback about whether their choice was right or wrong about the correct response and finally for the auditory-only condition group.

It has been demonstrated that the head and body dynamics strongly influence the spatial auditory perception, especially when the duration of the sound stimulus exceeds 250 ms [54]. Thus, the modifications occurring in the values of the binaural cues conduct to a better localization performance that is reflected in a lower incidence of front-back confusions [27].

Zahorik's experiment [55] presents a sound localization training procedure that offers the subjects auditory, visual and proprioceptive/vestibular feedback concerning the correct direction of the sound source in space in a virtual auditory environment where the main auditory stimuli have been represented by 3D sounds synthesized with non-individualized HRTFs. During the training session, the visual indicator of the correct position of the sound source has been paired with a repetition of the auditory stimulus in order to recalibrate both the visual and the hearing senses. All in all, the results of this experiment demonstrate that the listeners can adapt to altered listening conditions, such as the use of 3D sounds synthesized with non-individualized HRTFs, even after a short perceptual feedback based training session of approximately 30 minutes. Moreover, the improvements demonstrated by this study seem to be long-lasting (even after a time period of 4 months) and do not interfere or alter the spatial map representation employed by the auditory system under real-world listening conditions, supporting Hofman's argument that training conducts to the development of a secondary spatial map in the brain [35], [55].

In Carlile's experiment [50] into the ears of four groups of subjects have been inserted pieces of mold kept there for a period of 10 days. The control group did not receive any type of feedback during the training session, the second group received only visual feedback through a LED illuminated on the active loudspeaker, the third group was offered audio and visual feedback, while the fourth group benefited from audio and visual training with the room lights turned on during training. In the 10th day of training, the highest improvements have been recorded for the front-back confusion rate, particularly for the group which received visual and auditory feedback (the third group).

INTERACTIVE PERCEPTUAL LEARNING

In the "active training" paradigm [41] the subjects are actively engaged in the process of improving their spatial auditory localization skills. The training procedure does not have a fixed, pre-planned scenario and organization, while the feedback that is provided to the subjects results from their own actions and interactions with the virtual environment. For instance, Parseihian and Katz's study [42] presents a report on the degree of adaptation of a group of sighted subjects whose sound localization abilities have been trained in a virtual auditory environment with audio and kinesthetic perceptual feedback. The results of the experiment showed that both the angular precision errors and the front-back confusion rates reduced in the post-test session. Thus, the front-back confusion rate decreased from approximately 25-27% in the pre-test session to 11-18% in the post-test phase. The authors concluded that the sound localization improvements reported in the post-test session have been triggered by the process of adaptation of the human auditory system to the perception of 3D sounds synthesized with non-individualized HRTFs and are not the aftereffects of interactive gameplaying.

In the experiment conducted by Mendonça et al. [56] the sound localization performance of 12 inexperienced subjects in the horizontal and vertical plane has been tested before and after a series of training procedures. The results demonstrate that all the subjects succeeded in improving their sound localization performance. Also, the white noise offered a higher spatial auditory resolution, as the subjects reached a mean angular localization error of 13.8 degrees when listening to white noise, while the group who trained with speech stimuli obtained a localization error of 18 degrees [9], [10].

CHAPTER 3
SPATIAL AUDITORY REPRESENTATION IN THE CASE OF THE VISUALLY IMPAIRED PEOPLE

The blind persons are dependent on the auditory cues, since they allow them to identify the attributes of the surroundings, to recognize familiar or unfamiliar objects and to effectively manage the spatial information. Over the years, it has been argued that the visually impaired people benefit from augmented hearing and spatial localization abilities, due to sensory compensation and crossmodal plasticity [57]. Several studies have demonstrated higher sound localization skills in the case of early-blind individuals, others have concluded that there is no difference between the blind and the sighted group, emphasizing that vision is not responsible for enhanced spatial auditory perception, while others proved that the visually impaired listeners are less proficient than the sighted individuals in what concerns sound localization and spatial discrimination [58].

The state of the art presented in this chapter has been published in [58] and [59].

3.1. VISION AND SOUND LOCALIZATION

There are many experiments that studied the relationship between the visual and the auditory senses in respect to sound localization [60]. They have argued that the visual cues improve the sound localization performance, especially when the inputs from both senses are correlated [61], as when the subject listens to a target sound and concurrently visualizes its source. There have been several studies that demonstrated the effect of vision on training the sound localization abilities of sighted individuals [62]. The visual information played a fundamental role, as it helped the subjects to dynamically process the perceived sounds through a crossmodal sensory association. In consequence, vision calibrates the auditory perception, directs the attention towards the incoming stimuli and ensures the development of a solid representation of the environment. However, there should be made a clear distinction between the way in which the visual and the auditory channels encode the information received from the environment. For instance, the hearing sense uses the craniocentric frame of reference that depends on the position, direction and anatomical features of the external part of the auditory system (the pinna, which is well-known for its auditory filtering effects and the head, which causes attenuations and a head shadowing effect to the sound), while, on the other hand, the visual sense relies on the oculocentric frame of reference, which shifts focus with head movements [60].

In the case of the visually impaired people, the spatial representation of the environment is supported by another type of sensorial and perceptive organization. In the lack of sight, they are forced to use an alternate sensory modality, for instance hearing, to compensate for their visual deficit. The blind individuals rely intensively on the auditory sense in order to spatially map the external world [63] and to perform simple navigational tasks [64]. Auditory perception, as the inner process of unconsciously transforming the sound energy arriving at the eardrums into relevant information for the listener is significantly influenced by the characteristics of the incoming sound stimuli, generally referred to as “the auditory dimensions”. The auditory dimensions of sound (pitch, loudness, timbre, directionality, distance, externalization) convey important information concerning the shape, size and the other characteristics of the environment. All these cues are dynamically synthesized in a single, unitary image of space that is called “spatial cognitive map” [58], [59].

3.2. CROSSMODAL PLASTICITY IN THE LACK OF SIGHT

The information from the central and from the peripheral regions of space is processed differently by the auditory system of the visually impaired people. The human mechanism of spatial auditory processing is divided into the spatial “where” (manipulated in the posterior-dorsal region of the auditory cortex) and the spectro-temporal “what” (that is activated in the anterior-ventral pathway) streams, similar to the perception acquired by the visual system [63]. It has been demonstrated that the blind people are more proficient at detecting the origin and direction of a sound source (by activating the “where” pathway) than at recognizing the spectral and temporal features of the sound (in the “what” cortical area). For instance, in the sound localization experiment described by [65] that involved the participation of both sighted and visually impaired subjects, the latter group outperformed their sighted counterparts, especially at sound localization tasks in the peripheral region [65]. Moreover, in the same experiment, the neuroimaging analyses concluded that the posterior-dorsal auditory “where” pathway was more activated in the case of the blind participants. In the experiment presented in [63], several congenitally blind and sighted (blindfolded) subjects were required to undergo spatial localization and non-spatial frequency discrimination tasks under free-field listening conditions in the peripheral region of the space. Each trial consisted in the presentation of a sound cue over the left or the right speaker (a pure tone, 555 Hz frequency), followed by the onset of an additional sound at the same or at a different location, with equal or unequal frequency level. The subjects were required to locate the target sound and to specify if there is any difference in the spectral content of the two consecutive auditory cues. The results of the study concluded that the visually impaired subjects were able to detect and localize the peripheral sound targets more accurately than the normally sighted participants, although they were less efficient in the frequency discrimination task. These results provided supplementary evidence in favor of the previous findings regarding the improved functionality of the “where” pathway that helps the visually impaired people to focus on the most important features of the environment and to navigate more safely by avoiding dangerous situations [63].

Concerning the modifications that take place at the brain level, it has been observed (in the case of early-blind cats) that the occipital visual area has adopted some of the sound processing functionalities from the auditory cortex and at the same time the spatial auditory representation became more accurate [66], [67], [68]. The sound localization experiment performed by Gougoux et al [66] supported the hypothesis that auditory processing expanded into the visual cortex in the case of the early-blind subjects. This crossmodal plasticity was observed to occur especially under monaural listening conditions. The authors argued that the enhanced abilities for discriminating monaural sound cues are due to the recruitment of the occipital cortex for processing spectral cues. Another study emphasized that the early and late blind individuals recorded significantly better results than the sighted subjects after performing binaural sound localization tasks [69]. The level of plasticity of the auditory system varies among the blind listeners, depending preponderantly on the level of adaptation to the sound localization tasks. The fMRI studies performed by Collignon et al [70] demonstrated that the spatial-auditory processing mechanism of the congenitally blind people uses the regions of the dorsal occipital stream which are responsible for spatial visualization in the case of the sighted individuals. Thus, the spatial auditory localization of the congenitally blind people takes place in the same cortical regions that process visual inputs for the normally sighted persons. Moreover, the authors argue that this crossmodal plasticity is unselective and does not require experience or practice to ensure a high degree of sound localization specialization in the occipital cortex.

According to Chan et al [71], the parietal cortex plays an important role in creating sound-to-distance associations in the case of the blind people. Furthermore, the hippocampus is implicated in the spatial process of dynamically linking sounds and locations, demonstrating that the crossmodal plasticity involves other cortical regions of the brain, besides the occipital cortex (Fig. 3.1.). The study concluded that auditory localization in the case of the blind individuals requires the participation of the visual cortex and of the hippocampus more actively than in the case of normally sighted subjects (who rely on the activity of the frontal and temporal lobes).

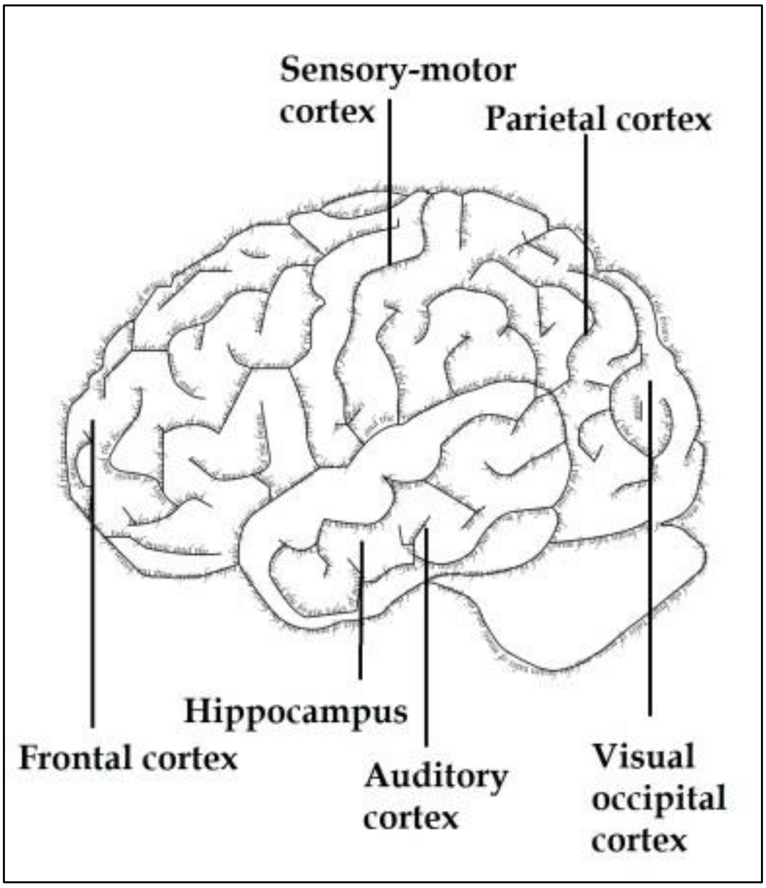

Fig. 3.1. The sensory regions of the brain

Besides crossmodal plasticity, intramodal plasticity also influences sound localization accuracy. There have been studies that showed a certain level of intramodal plasticity in the auditory cortex, particularly in the tonotopic area of the primary auditory cortex (A1). The tonotopic area, which is responsible for decoding frequency, proved to be significantly enlarged in the case of the blind people [58], [72].

3.3. TRAINING AND BRAIN NEUROPLASTICITY

Even in the case of congenitally blind people, who do not possess any sight experiences or memories at all, the neuroplastic process of adaptation and specialization of the occipital visual cortex can be enhanced as a result of training, due to the inborn abilities of the brain.

Training plays an important role in improving the cognitive abilities of the blind listeners. The subjects who received intensive auditory training were able to better process the information and to assess the visual principles of navigation and orientation [73]. Essentially, blind people use primarily an egocentric system of reference, adjusting the overall perception of space according to their own position. The experiment of Iachini et al [133] demonstrated that for the congenitally blind people it was more difficult to represent the space allocentrically, in comparison to the late-blind and sighted subjects, particularly in what concerns the large-scale spaces. On the other hand, the egocentric, body-based representation was better for all the groups in a small-scale shape. In addition to this, training is more efficient if it is focused on the needs of the user. For example, an effective practice should concentrate on the employment of dynamic and interactive lessons and on the use of functional tools such as virtual environments or games [58].

3.4. COMPARATIVE SOUND LOCALIZATION EXPERIMENTS WITH NORMALLY SIGHTED AND VISUALLY IMPAIRED SUBJECTS

This subchapter aims to present the most relevant sound localization experiments that involved the participation of both normally sighted and blind subjects. As these studies provided different conclusions, we have classified them into experiments that present either better or worse localization performances in the case of blind individuals or equal accuracy for the two groups.

3.4.1. Experiments demonstrating improved sound localization performance in the case of blind individuals

The results of the tests that were performed under both monaural and binaural conditions showed that the congenitally blind subjects can localize sounds with equal or better accuracy than the sighted individuals. Furthermore, the subjects with residual vision were less accurate than the normally sighted participants, demonstrating that crossmodal compensation develops in time, as it may be correlated with the patient's medical condition and with the duration of sight deprivation [74].

This argument is also supported by the study of Doucet et al [75], who investigated the binaural and monaural localization accuracy of blind and sighted control subjects in the horizontal plane. By contrast to the sighted subjects, half of the blind participants showed enhanced auditory discrimination abilities under monaural listening conditions, having one ear blocked. Their performance was consequently tested by either obstructing the pinna and by using low-passed and high-passed signals with the ears unobstructed. The results of the tests demonstrated that the visually impaired individuals possess an enhanced capacity (considered by the authors as supra-normal) of discriminating and processing the spectral content of the sound.

The results of Ohuchi et al [76] demonstrated that the blind subjects outperformed the normally sighted participants under both head-fixed and head-rotating listening conditions and that the localization errors when the subjects were allowed to rotate their heads were slightly lower than in the head-fixed condition. Moreover, the blind listeners were able to accurately report the distance to the sound sources (especially when they kept their heads fixed), by comparison to the sighted participants, who showed a clear tendency of underestimating it.

The results of the experiment described in [65] indicated that the blind subjects were more proficient at performing sound localization tasks in the peripheral region of space than the normally sighted individuals. This particularity was observed in the areas where localization accuracy is most inaccurate for the sighted people, such as the lateral directions. Also, the authors suggest that this enhanced sound localization performance in the peripheral space is influenced by the selective attention process that occurs within the first 100 ms after the sound onset.

A similar experiment that investigated the ability of congenitally blind and sighted individuals to perform spatial (detection and localization) and non-spatial tasks (recognition of spectro-temporal auditory features) in the peripheral auditory region concluded that the blind subjects were able to localize the sound sources more accurately than their sighted counterparts by showing the same degree of selective attention. Nonetheless, the congenitally blind participants were slower at discriminating the frequency components of the sound, demonstrating that the "what" detection mechanism of the auditory system is less specialized than the "where" localization process. As discussed earlier in this chapter, this behavior is highly ecological, since it supports orientation and cardinality for the visually impaired people.

In the study of Wersényi [77], the sound localization performance of 28 blind and 40 sighted subjects was tested in a virtual auditory environment. Briefly, the results of the experiments showed that the blind subjects reported better results for sound localization in the 3x3 grid virtual display and for detecting the direction of the static source, especially in the front. In what concerns the dynamic sources, they were able to identify the direction of movement of the secondary target source more accurately in the horizontal plane. Furthermore, the blind listeners showed a higher spatial discrimination ability during the Minimum-Audible-Angle measurement, outperforming the sighted group of users by 1-2 degrees.

The results of Voss's study [69] demonstrated that the early-blind group outperformed the sighted and the congenitally blind groups when the sound was emulated in the frontal hemifield. When the sound was presented behind the interaural plane, the results of the early and late blind groups were indistinguishable, but significantly better than those of the normally sighted control group. Thirdly, the results of the minimum-audible-distance task reported that the blind listeners were more proficient than the sighted subjects at differentiating the distance between consecutive audio stimuli under free-field listening conditions [58].

3.4.2. Experiments demonstrating equal or lower localization performance between the sighted and the visually impaired people

Zwiers et al [78] performed a series of sound localization experiments that studied the auditory localization accuracy of two groups of sighted and early-blind subjects. The results demonstrated that the azimuth localization accuracy of the blind subjects when they were exposed to long-duration, broadband Gaussian white noise stimuli was comparable to that of the sighted individuals. Besides, the elevation localization accuracy was also indistinguishable between the two groups of subjects, independent of the pointing method that has been employed.

In the experiment described in [79], the sound localization performance of 5 congenitally blind and 5 sighted subjects was compared and discussed. The listeners were required to turn their heads towards the perceived direction of the incoming stimuli. The results of the experiments demonstrated that the sound localization performance of the blind subjects was not as accurate as that of the sighted individuals.

Another experiment that demonstrated a decreased spatial auditory resolution in the case of the blind people is presented in [80]. The sound localization procedure took place under virtual auditory conditions and the subjects were required to listen to 3D sound stimuli. The results revealed higher localization errors in the case of the blind subjects, with an increased level of inaccuracy in the vertical plane.

Katz and Picinali [81] designed a large scale immersive virtual auditory environment consisting of a room (presented both physically and virtually) where six sound sources were arranged around the listener. The stimuli were very well-known auditory cues, such as the sound of running water, telephone ringing, dripping faucet, washing machine, coffee machine and a ticking clock. The 54 subjects who participated in the experiments were divided into 3 groups: congenitally blind, late blind and sighted (blindfolded) subjects. The results reported no significant difference between the blind and the normally sighted group in what concerns the absolute distance error (the difference between the correct and the perceived location of the sound source). In respect with the angular error, the congenitally blind people recorded worse results than the late-blind and sighted subjects.

3.4.3. Conclusions

The results of the previously presented experiments demonstrated that the blind individuals possess enhanced sound localization abilities in the horizontal plane (while the sighted subjects are more proficient for vertical sound discrimination) and for performing spatial localization tasks under monaural listening conditions. Moreover, the blind subjects are more accurate at differentiating the direction of the sound targets in the peripheral region of the auditory space, as this behavior is vital for adaptation and navigation in real-world situations. It has been also demonstrated that the blind individuals are able to identify more quickly the direction of the sound than its spectral features. This is a highly ecological consequence of the blind condition, as for the visually-deprived people it is far more important to react to the location of the stimuli (for instance, the sound of a car approaching) than to distinguish its spectral shape and characteristics [65]. On the other hand, there have been several studies that reported equal or reduced localization performance in the case of the blind individuals. Thus, it has been argued by some researchers that the lack of visual stimuli does not influence the spatial auditory resolution. By contrast, other authors consider that the visual cues are a fundamental prerequisite of the human auditory system for properly mapping and recalibrating the representation of the extra-personal space [58].

CHAPTER 4
AUDIO GAMES FOR THE VISUALLY IMPAIRED PEOPLE

This chapter presents a review of the audio games for the visually impaired people. The contents from this chapter have been published in [82], [83] and [84].

4.1. GENERAL ASPECTS OF AUDIO GAMES

As the majority of computer games available on the market today are based on graphical user interfaces, the visually impaired people have constrained access to the use of this important means of interaction they could greatly benefit. The rapid increase in consumer multimedia technology and the need for an accessible source of entertainment for the blind people has led to the evolution of a new game category, i.e. the audio games.

4.1.1. Game accessibility to the blind people

As most of the computer games available on the market today are based on visual stimuli, the people who suffer from sight disabilities are facing serious user interaction limitations. To overcome this drawback, many video games contain accessibility facilities, such as high-contrast graphics, scalable text and custom color palette settings for the low-vision or colorblind players [85]. Nonetheless, the users who are totally blind cannot benefit from these aiding features and consequently find themselves in the impossibility of playing any computer game.

4.1.2. Types of audio games

In audio games, the main sensory channel used for the transfer of information from the computer to the end-users is hearing. According to one simple classification, audio games are divided into 2 categories: games that are exclusively audio, called audio-only games - designed for the blind people and on the other hand, hybrid games that combine both graphical and auditory features for the low-vision players. Audio games are usually developed by small companies with little funding, research communities or by individuals who suffered from visual problems themselves. Nonetheless, they are addressed to a wide range of public, being accessible to both blind and non-blind users. The purpose of the audio games is to express all the information through sound and music in order to create an immersive virtual environment. Over time, the audio games have been released as PC games, online and mobile software and applications for handheld devices. The focus is concentrated on the sense of hearing and on the players' capacity to perceive, identify and discern between various sound cues [82].

4.1.3. Benefits of audio games

The audio games improve informational and contextual learning and situational knowledge. Also, these games play a significant role in the formation of behavioral abilities and mental structures by enhancing the player's problem-solving skills, by improving creativity, imagination and by stimulating learning. In addition to this, navigational audio games promote a solid mental representation of the environment, a higher level of independence and spatial reasoning, robust exploratory abilities and strong searching and way-finding capabilities [82].

The results of the experiment described in [86] demonstrate that the audio game performance can be transferred in real-life navigation situations. These findings are remarkable, especially because good results have been observed in the case of the people who have never undergone any orientation or mobility training before. As a result, audio navigational games serve not only as a means of training the spatial representation skills of the visually impaired people, but also as an effective way of translating the acquired abilities into the fulfillment of real-world exploratory goals. Moreover, an experiment in which the players were required to play the Blind Side game demonstrated that the blind people succeeded to finish the game earlier than the sighted subjects [82], [84].

4.1.4. Sense of presence and flow

Audio games offer a wide range of emotions that are incorporated into 2 main concepts: presence and state of flow [87]. The feeling of presence is defined as the illusion of being part of the game, while the concept of flow refers to the sensation of pleasure, absorption and enjoyment that the player experiences as a result of a challenging ludic activity.

Flow, as described in [88], represents the feeling of "happiness, creativity, subjective well-being and fun" that people encounter when they are involved in an activity which brings them a high level of excitement. Nowadays, video games generate flow experiences by promoting positive feelings, entertainment, immersion and a high degree of motivation. In order to be appropriate for the blind people, the audio games should draw a balance between the user's abilities and the scenario's challenges. Moreover, they need to fulfill the following 7 requirements suggested by Csikszentmihalyi [89], the psychologist who identified the concept of flow in the mid-1970s [88], [89]:

- A demanding activity that intensively requires the abilities of the player;
- Well-defined, comprehensible goals;
- Concentration on the current task of the game;
- The sensation that the player loses track of time while being immersed in the game;
- A feeling of being in control of the game's outcome at each moment of time;
- The sense of immersion and loss of self-consciousness;
- The user should be immersed in the action of the game and concurrently be aware of the changes that take place at each moment of time.

4.1.5. Recommended features of audio games

Blind people should perceive audio games as a toy - an easy and interactive method of learning and a reliable source of entertainment at the same time. To be accessible and appealing to the visually impaired people, the audio games should fulfill the following specifications:

- A high level of immersion and attractiveness, in order to determine the user to advance in the game;
- A diversity of audio cues;
- A small amount or a complete lack of "game over" situations that would discourage the user from continuing playing the game;
- A single, clear way of resolving the issues raised by the scenario;

- An intuitive learning strategy that should be integrated in the outline of the game. In this way, the visually impaired people understand the game's requirements without the use of human or text-to-speech instructions;
- Interactivity, engagement and fun, in order to make the blind people enjoy a pleasant experience [90].

4.2. SONIFICATION IN AUDIO GAMES

Sound plays a very important role in computer games, as it makes the player feel immersed into the action, giving him situational knowledge and relevant information about the presence of objects and characters. Also, it contains emotional information, "triggering feelings and memories" [91].

In audio-based software applications, such as auditory displays and audio games, sonification is used to represent various actions, objects and situations and to describe the storyline. In this way, the whole picture of the game is translated into auditory stimuli, making the gameplay accessible to the visually impaired people.

The most important elements of sonification are:

- Auditory icons - sounds that create an analogy with real-world events and situations. They represent real world events, such as the sound of a door opening, footsteps on the ground, a gunshot or a clock ticking.
- Earcons - abstract, symbolic sounds (noise, artificial sounds) used to facilitate the players' displacement through the complex content of the game. They are required for the transmission of short messages that do not demand advanced cognitive abilities. A good example of earcons is the sound of an emerging alarm that represents danger.
- Speech - an aiding sonification technique which helps the players to overcome difficult situations that cannot be resolved without verbal instructions. Narrators are used to guide the players (for example, in exploration games, they give information regarding the current position of the avatar, assign challenges and drive the plot's outcome in a natural way).
- The monologues of the main character – they simplify the scenario and give clues about goals, objects and actions. Their purpose is to assist the players and to make them acquire a general perspective of the game's scene [92].

An essential aspect that has to be taken into account when designing audio games is the fact the developer has to draw a balance between the number of audio sources (playing simultaneously or over time) and the information conveyed by them. Also, the user's immersion can be improved by using an exaggeration of special audio effects or other non-realistic, artificial sounds.

3D sound is very effective for transferring directional information, especially in navigational games. Music and ambient sounds establish a general background atmosphere, communicate information about the emotional aspects of the game (funny, pleasant, intriguing, action-based or horror situations) and contribute to the temporal rhythmic progression of the action (slow, moderate or fast).

Audio games have a double nature: they are primarily games designed with the purpose of improving user interaction and, on the other hand, they can be regarded as fictional, imaginary worlds defined by a high level of immersion and an acute sense of presence. Sound is used to highlight the features of the fictional world depicted in the game and to enhance drama and immersion [83], [93].

4.3. CATEGORIES OF AUDIO GAMES

Audio games are divided into various groups, according to their interaction approach. The main types of audio games are: action, adventure, puzzle and strategy games.

Action games require good attention and synchronization from the user. The timing of reaction and interaction is very important because the outcome of the game is strictly related to the player's quick feedback.

Adventure games have the following characteristics: an exciting plot, activities based on navigation and exploration and the existence of a hidden mystery that needs to be uncovered. A relevant example of narrative adventure game is The Hidden Secret [94].

Strategy games require the player to manage different situations and to manipulate the given resources. Puzzle games enhance the users' memory and attention and stimulate the cognitive brain processes. For instance, the jigsaw audio-based puzzle described in [95] is an interactive approach towards entertaining oneself with music in a very exciting way. The purpose of this game is to rebuild a fragmented song from random audio pieces.

Another category of audio games are the sports-oriented games. For instance, Audio Soccer [96] is a relevant example of sports-based game made accessible to the visually impaired people through sound. It employs 3D sound that defines the movements of the ball such as passing and tackling, the user focus, the playing settings (the borders of the pitch), and the current state of the game. The player's orientation is provided in response to the sounds produced by the other players (echolocation) [82], [84].

Slingshot 3D [134] is an interactive haptic-audio-video shooter game that offers two types of haptic feedback: tactile feedback, through a haptic jacket and kinesthetic feedback, through the Novint Falcon interface. An experimental procedure showed that the players enhanced their immersion and gameplaying performances as a result of the haptic feedback provided by the game.

Audiopolis [135], a game that is played from the first person perspective, helps the blind players to navigate in real and virtual environments. It has three interfaces: haptic, audio and haptic plus audio. The Novint Falcon device offers force-feedback information that can be obtained by touch (using the hands or a white cane), concerning the physical characteristics of the surrounding objects. In the environment, the user can navigate, explore different spaces, identify objects and advance to the further levels.

4.4. NAVIGATIONAL AUDIO GAMES

In the case of the blind individuals, the development of navigational skills is based on the remaining sensory modalities, such as hearing and touch. As the tactile sense is limited to the perception of nearby objects, hearing is a more powerful sensory channel, due to its large capacity of perceptual input. As it is generally believed that the deprivation of sight is compensated by a behavioral improvement of the remaining senses, the audio games prove to be an efficient rehabilitative technique towards training and testing the navigational skills of the visually impaired subjects within an immersive and user-centered virtual reality environment.

Navigational audio-based games enhance the development of spatial cognitive skills and help the blind people to access and manipulate environmental information by stimulating contextual learning. Moreover, contextual learning can be successfully transferred to the real world, raising sensory awareness, promoting searching skills and improving the localization abilities in unknown environments. All the games presented in this subsection have been designed from the “first person” perspective, except for Terraformers [100], which uses the “from above” player approach.

AbES (Auditory based Environment Simulator) [86], [97], [98] is an audio game that simulates navigation inside a virtual building, starting from the premise that the acquired spatial contextual learning can be transferred into real-world exploration.

Virtual environment: A virtual rendering of a physical building (a 2-story building with 23 rooms, two stairways and 3 exits) that contains structural elements such as walls, doors, rooms, corridors and other furniture objects - doors, desks or tables.

Game scenario: The purpose of the game is to search for randomly hidden jewels and to bring them out of the building as quickly as possible, by avoiding monsters that try to take and hide them in another place.

User interface: Audio and graphic interface; the audio interface helps the user to identify his position, orientation, heading, as well as the location of nearby objects; the graphical interface gives clues about the layout of the building, the current state of the game and the position of the objects and the avatar.

User interaction: Keyboard interaction: space bar (go forward), H key (turn left), K key (turn right), J key (action - for example, open a door), F key (help key - information about the current position of the player).

Sonification approach: 3D naturalistic and comprehensible spatialized sounds; auditory icons; earcons. Text-to Speech soundscapes - audio verbal commands; orientation (north, south, east, west), location and heading are based on TTS (Text-to-Speech) commands; distance is encoded by the sound intensity and elevation by the pitch variance.

Transfer of learning: The transfer of learning has been demonstrated by an experiment that offered significant results, suggesting that the players were able to transfer the acquired spatial knowledge into real-world situations.

The TIM project (Tactile Interactive Multimedia) [99], was initiated by SITREC (Stockholm International Toy Research Center). Tim’s Journey is an adventure game that brings together complex and interactive soundscapes, along with a “non-linear narration and open-ended gameplay” [99].

Virtual environment: The game’s environment is an island which contains different scenes - a harbor, a forest, a mill.

Game scenario: The player moves around freely in order to discover a hidden mystery.

User interface: Both audio and graphical interfaces; proper user navigation is assured by: the ambience reductor (with the purpose of lowering the volume of all the objects that the user cannot interact with), footsteps (give information about the travel surface), helpers (characters that offer information), foghorns (indicate the direction on the compass).

User interaction: Keyboard.

Sonification approach: 3D directional sounds; each scene presents a different musical theme; the musical pieces are changing continuously, as they are the result of the sounds corresponding to each action in the environment; the importance of the objects is emphasized by the intensity and repetition of their associated soundscapes.

Audio Doom [99] is the auditory-based correspondent of the exploratory game called “Doom”. The aim of Audio Doom is to help the blind children to perform navigational tasks in order to improve their spatial cognitive abilities, orientation, mobility and problem-solving skills.

Virtual environment: A 3D labyrinth of walls and corridors.

Game scenario: The player explores the virtual environment while trying to avoid monsters and reach the exit towards the next level.

User interface: Graphical and audio interfaces that allow the players to sequentially navigate through the game settings.

User interaction: Keyboard, mouse, joystick.

Sonification approach: Audio spectral cues (for example, the sound of a door opening or footsteps).

Transfer of learning: An experiment with 7 blind children demonstrated that by playing the Audio Doom game the subjects can improve their spatial cognitive skills and alternatively develop good tactile representation abilities. For instance, the participants in this experiment were able to rebuild the route they traveled in the game using Lego blocks.

Terraformers [100], released in 2003, is an audio game accessible to both sighted and unsighted players.

Virtual environment: A 3D virtual environment in which the action takes place in the extraterrestrial space, on the imaginary planet Tellus 2.

Game scenario: The mission of the game is to defeat the revolting robots that colonize the planet, to recover the computer components that they have scattered and to free their creator, Prof. Lange.

User interface: Both visual and audio interfaces; a high contrast mode that presents parts of the 3D graphics in black and white in order to be accessible to the low-vision subjects.

User interaction: Keyboard.

Sonification approach: 3D sound is assigned to all the game's objects; voice feedback; sound compass - voice command that indicates the direction (north, south, east, west); sonar - 3D sounds that indicate the distance to target objects; GPS position system that indicates the current location of the avatar though speech commands; ambient environmental audio cues.

The game strategy presented in **Pyvox** [90] intends to create an enjoyable experience, while maintaining a balance between the challenge of the game and the skills of the players(the flow theory).

Virtual environment: A 70-floor labyrinth tower that corresponds to 70 different levels of the game; each level has an exit and a number of walls.

Game scenario: The game's goal is to find the exit of each floor, without hitting the walls. When the task is accomplished on one level, the player can ascend to the next one.

User interface: Both audio and visual interfaces.

User interaction: Mouse cursor movement.

Sonification approach: No verbal instructions at all; the exits on one level are indicated by a stereo rendering that presents a variation of pitch according to the current position of the player in respect to the location of the exit.

Transfer of learning: An experiment performed with 2 groups of blind users demonstrated that all the players have well understood the rules of the game and managed to pass the first three levels.

Blind Side [101], [102], [103] is an audio game developed by Aaron Rasmussen and Michael Astolfi whose scenario has been inspired by Aaron's personal experience of being temporary blind after a chemistry accident that took place during high school. Blind Side aims to convey a realistic and immersive audio experience that transmits the emotions of rediscovering the world after waking up completely blind.

Virtual environment: An audio-only virtual environment representing the settings of a building and a city.

Game scenario: An assistant professor called Case wakes up totally blind near his girlfriend, Dawn. The goal of the game is to help Case and Dawn to navigate through the environment and to rediscover the world in their new condition of blind people.

User interface: Audio interface only; available on Mac, PC and for the iOS mobile devices.

User interaction: Keyboard; gyroscopic control scheme.

Sonification approach: 3D audio; over 1000 recorded, realistic sounds; narration system; low-pass filtering and attenuation for the sounds corresponding to the locations situated behind the listener; dynamic reverb; trailing effect that simulates the contact with the surrounding objects - the sound varies in pitch according to the angle and speed the subjects are sliding along the surface of the objects.

Transfer of learning: An experiment that took place at Smith-Kettlewell Eye Research Institute demonstrated that the blind people usually finish the game earlier than the sighted players [82], [84].

4.5. CONCLUSIONS

The audio-based games presented in the previous subchapter provide different user interaction approaches towards navigating in a virtual auditory environment. As either the plot involves uncovering target sound emitting sources, avoiding obstacles or finding the exit from a building by navigating through a series of rooms and corridors, these audio games require intensive attention and concentration to the perceived auditory stimuli. The sonification approach in these games is based on 3D directional sounds, realistic recorded sounds, auditory icons, earcons and narrative speech that gives advisory information on the current state of the game or introductory clues concerning the manipulation of the user interface menu. 3D sound enhances situational awareness by providing directional clues, enriches the amount of information conveyed through auditory stimuli and compliments the 3D perception by extending it on the auditory level [104]. Moreover, the dimensions of sound (amplitude, pitch) are altered in order to convey information regarding the current position of the objects and the relationship between the player and the targets (distance, azimuth, elevation). Nonetheless, it is important to take into account the role of ambient sounds which create a powerful immersive effect that introduces the player into the atmosphere generated by the game's scenario. Moreover, the experiments conducted on the visually impaired people who played audio games revealed a significant improvement in the acquisition of cognitive behavioral skills, contextual learning and spatial environmental representation. These results can be transferred for performing real-world navigational tasks, fact that fulfills the objectives of the virtual auditory simulation.

The relevant aspects that need to be researched further are whether the virtual auditory environments can be broadened in design and functionality to support a wider range of problem-solving tasks [105]. In addition to this, the user interface needs to be thoroughly designed in order to become accessible and operative for the blind people [82], [83], [84], [106].

CHAPTER 5
APPLICATIONS DEVELOPED FOR IMPROVING THE SOUND LOCALIZATION PERFORMANCE, EXPERIMENTS AND RESULTS

This chapter presents a comparative study on the degree of spatial auditory improvement achieved by a group of sighed and visually impaired individuals who participated in two sound localization experiments that used a multimodal perceptual feedback based training procedure. The sound localization performance was assessed before (in the pre-test), after (in the post-test) and during the training sessions. The results demonstrated that both the sighted and the visually impaired subjects succeeded in improving their spatial auditory resolution, fact that has been reflected in a higher angular precision accuracy, a lower front-back confusion rate and enhanced navigational and orientation skills in a virtual auditory environment where the 3D sound stimuli have been synthesized using non-individualized HRTFs. The experimental procedures and results presented in this chapter have been published in [107], [108], [109], [110] and [111].

5.1. THE AUDITORY AND HAPTIC DEVICES USED IN EXPERIMENTS

5.1.1 The headphones

The auditory stimuli were delivered through a pair of stereophonic headphones (Sony MDRZX310L open headphones with no external correction of frequency characteristics). The presented level of the sounds was set to be confortable for the listener, having on average around 65-70 dB SPL [109].

5.1.2 The haptic belt

The haptic device that has been used in the experiment with the visually impaired people was designed by a team from the Technical University of Iasi [112]. It is easy to be used and can be adapted to research activities or real world scenarios.

The hardware platform has a 32 bit microcontroller, with superior energy consumption rates and math capabilities compared to 8 bit and 16 bit microcontrollers. In addition, the microcontroller has a Peripheral Reflex System (PRS), which enables the peripheral modules to communicate independently of the CPU (Central Processing Unit) and at low frequency rates, allowing it to save energy and to be kept at low power modes for a longer period of time. The system is battery powered, with a DC-DC converter that has input voltages ranging from 0.65 V to 5.5 V [112].

It is is composed of the following parts [107]:

- A USB-Wireless Gateway Device (UWGD) that allows the system to be controlled by the PC (it connects to the PC via USB).
- The Haptic Actuator Device (HAD), which controls the haptic actuators (Eccentric Rotating Mass – ERM motors) by receiving commands from the UWGD and executing them. The DRV2560L haptic driver uses a serial I2C interface that connects

the driver with the host microcontroller and a TCA9548A switch circuit with a one-to-eight architecture, meaning that up to eight I2C devices can be connected to the same bus. The current configuration of the device uses three I2C switches, resulting in a total of 24 vibrating motors [112]. Also, the haptic driver employs a licensed version of the TouchSense 2200 library [113] that integrates 100 haptic effects. Both the UWGD and the HAD use an aceMOTE v1.0 platform [114]. The control and communication unit is presented in Fig. 5.1.:

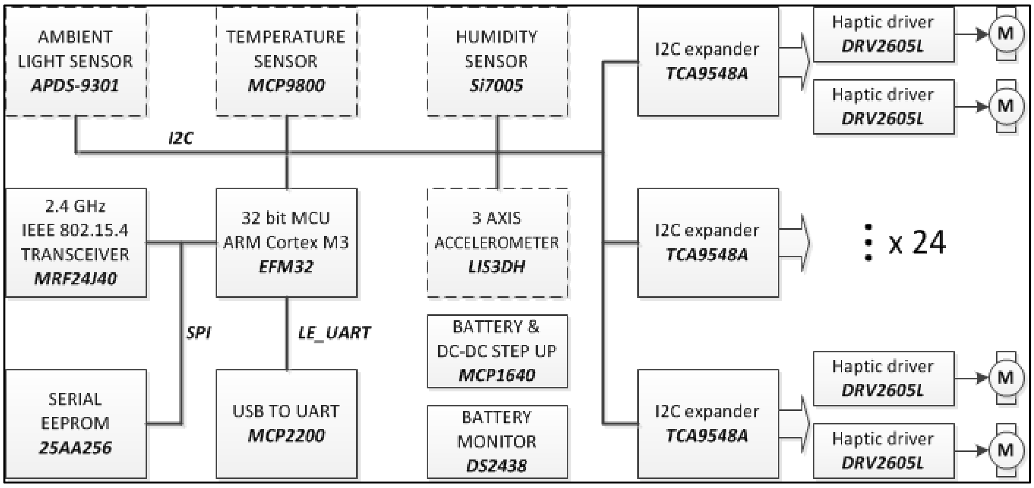

Fig. 5.1. The control and communication unit of the haptic belt [112]

- 24 vibration motors, of which only 12 have been used – each vibration motor is fixed along a stick in order to provide easy handling. The ERM haptic actuators rotate and the centripetal force causes the motors to move. As the number of rotations is high, the motors are displaced by the asymmetric forces, causing a repeated movement that is perceived as a vibration [112], [115]. The vibration motors have been placed at 30 degrees distance on the haptic belt, all around the listener's head. The structure of the haptic device is presented in Fig. 5.2. and Fig. 5.3 [110], [107].

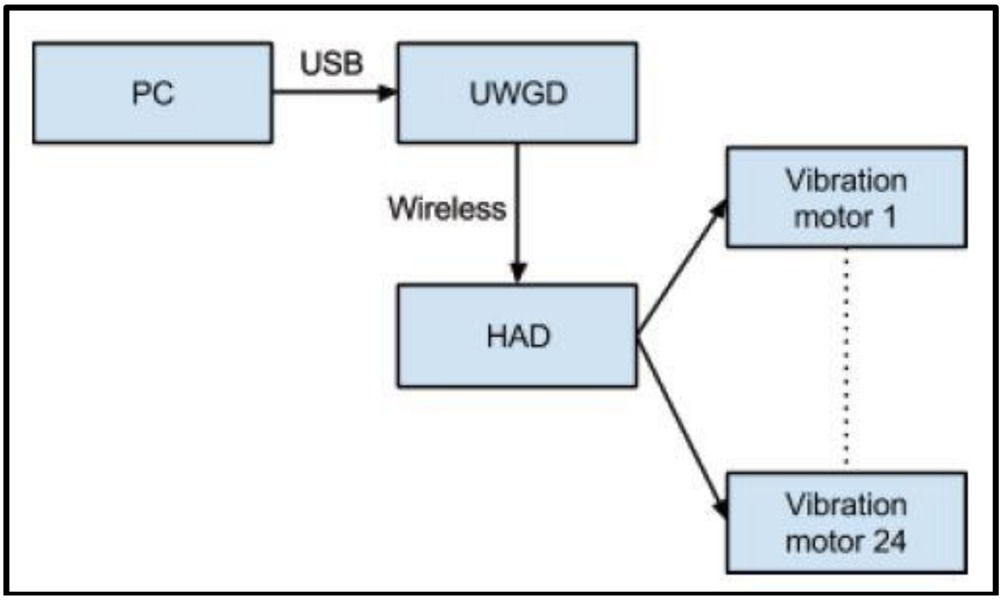

Fig. 5.2 The structure of the haptic device

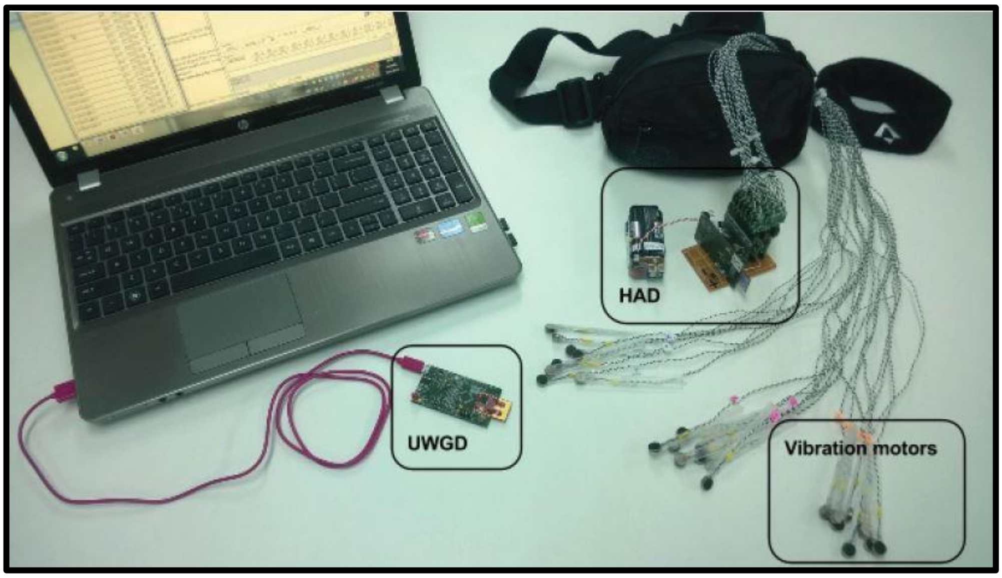

Fig. 5.3 The physical components of the haptic device

The haptic device interacts with the computer through a simple communication protocol. The developer is required to write the message format that makes the connection between the UWGD and the HAD device. This way, any programming language that supports serial communication (we used the C# programming language) can be used to develop applications for the haptic device [112].

In order to use the haptic belt from a PC application, the developer needs an USB 2.0 connection and the driver for the Microchip MCP2200 UART USB bridge that can be downloaded and installed from [116].

The command for the haptic device has six fields (Table 5.1.):

Table 5.1. Command fields of the haptic device

Field	Sample values (characters)
Command ID Always has the value 'S'. It represents the beginning of a command.	'S'
Motor ID A number between 1 and 24. It represents the vibration motor that must be actuated. It is ***ASCIIDEC*** encoded: number 10 is encoded as the string "10" (two characters: '1' and '0').	"01" for motor 1 "20" for motor 20
Vibration Type A number between 1 and 127. It represents the vibration type that will be produced by the motor ID. It is ***ASCIIHEX*** encoded: the decimal number 30 is encoded as the string "1E" (two characters: '1' and 'E').	"48" for the vibration type 4*16 + 8 = 72 "5A" for the vibration type 5*16 + 10 = 90
Number of repetitions A number between 1 and 4. It represents the number of repetitions that the motor will vibrate using the vibration type. It is ***ASCIIDEC*** encoded.	02" for two repetitions
Reserved field Unused. Reserved for future use. Must be "00"	00"
Command terminator One byte with value 0x0a. It represents the end of a command.	0x0a

A command example is "S01400100\x0a", meaning that vibration motor 01 is activated with vibration type 40 for 01 times.

5.2. SOUND STIMULI USED IN THE EXPERIMENTS

The sonification approach is aimed at reducing the incidence of reversal errors (front-back and back-front confusions) in the virtual auditory environment, using a method based on the spectral coloration of the sound which consists of listening simultaneously to two types of noises with different spectral characteristics (white and pink noise). The sound stimuli used in the experiments are a train of combined continuous Gaussian white and pink noises that were perceived simultaneously, but at varying auditory levels, according to the direction of the sound source in space.

At 0 degrees to the front the listeners could hear 100% white noise and 0% pink noise, consequently in the first quadrant the level of pink noise increased and the level of white noise decreased, reaching 50% white noise and 50% pink noise at 90 degrees (to the right side) and 0% white noise and 100% pink noise at 180 degrees (to the back). In the third and in the fourth quadrants, the rate of white noise increased and the percent of pink noise decreased, reaching 50% white noise and 50% pink noise at 270 degrees (to the left side) (Fig. 5.4.).

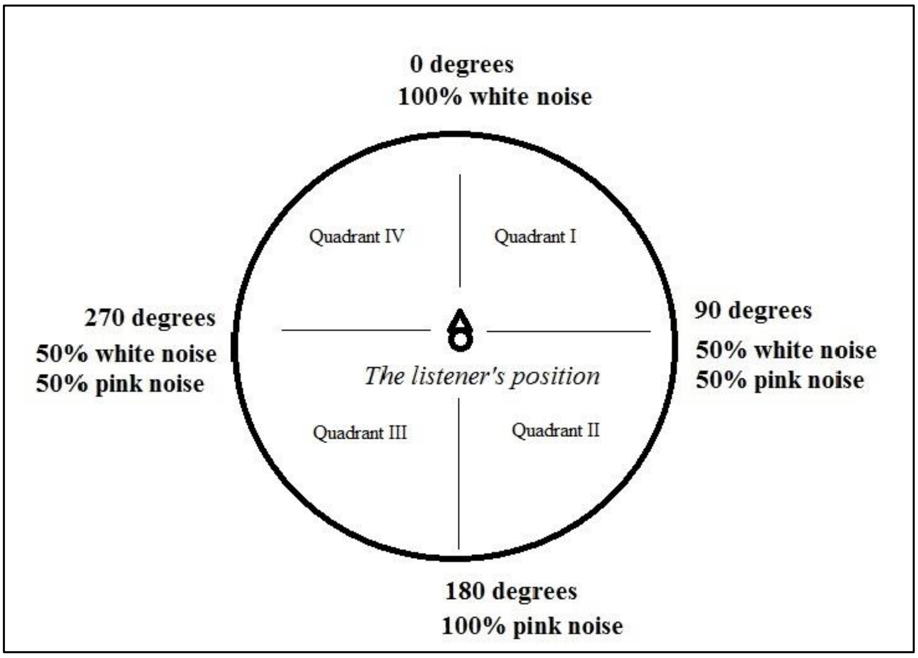

Fig. 5.4. The proportion of white and pink noise for the main directions (front, back, left and right)

The formula for calculating the perceived level of white and pink noise for a given angle in the horizontal plane is:

$$pink = angle / 180; \; white = 1 - pink; \; (0 \leq angle \leq 180)$$
$$white = (angle - 180) / 180; \; pink = 1 - white; \; (180 < angle < 360) \quad (5.1)$$

In addition to the white-pink noise combination, a discrete and repetitive (with 250 ms breaks between two consecutive bursts) “ding” type signal has been used in our experiments (Fig. 5.5.).

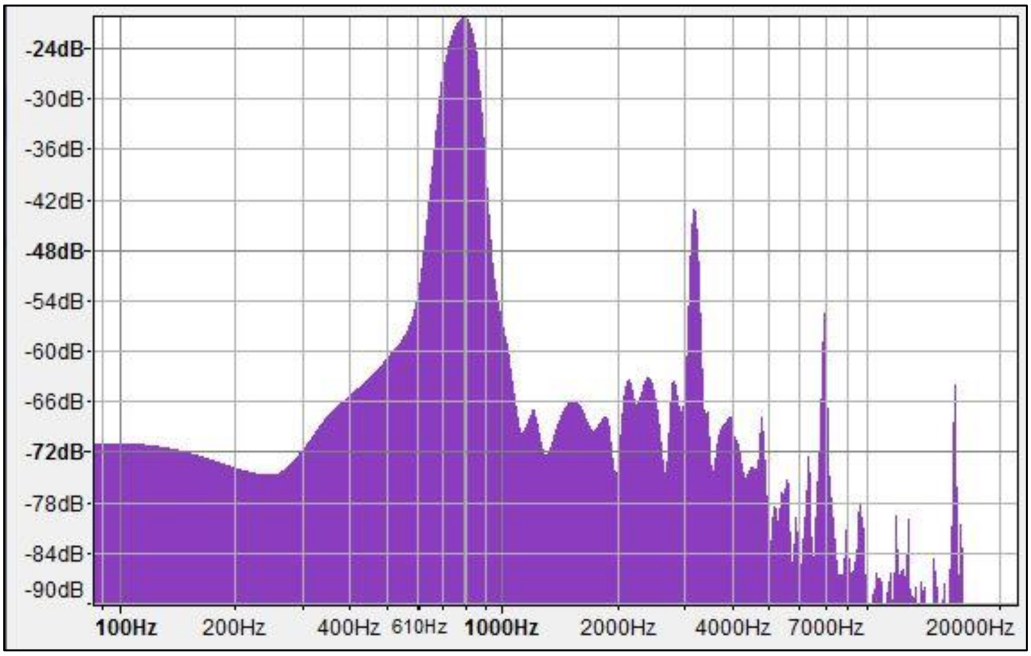

Fig. 5.5. The spectral profile of the "ding" sound with a peak at 1000 Hz

Both the white-pink noise combination and the "ding" sounds have been convolved in real time with the non-individualized HRTFs from the MIT database [117], using the CSound programming language for sound processing in order to obtain the 3D sounds that correspond to the desired angular directions in the horizontal plane [118].

The HRTFs from the MIT database have been recorded using a dummy-head mannequin and a set of 7 loudspeakers placed at 1.4 meters distance from it. The elevation positions in the median plane range from -40 degrees (below the interaural axis) to +90 degrees (directly overhead). The azimuth measurement locations were disposed at 5 degrees angle increment all around the listener's head [117].

Csound is the first sound processing programming language, developed in 1984 at the Massachusetts Institute of Technology and written entirely in C. Since then, it improved considerably, becoming today one of the most powerful and reliable music generating instruments. Moreover, it is available for the Linux, Macintosh and Windows operating systems, as well as for the new mobile platforms, such as Android [118].

In order to obtain 3D binaural stimuli for our game, we convolved the stereo sounds using the hrtfmove2 opcode (orchestra function, typical for the Csound programming language) [118]. The hrtfmove2 opcode generates 3D binaural sounds to be rendered over headphones for any location in the 3D space using the magnitude interpolation algorithm and a phase model based on the Woodworth's formula (equation 2.1.) [12] for Interaural Time Difference (ITD) [12], [27].

Magnitude interpolation means that the four nearest HRTF values (left, right, below and above) are linearly interpolated in order to obtain the most appropriate magnitude level for the given angular position [119]. The piece of code that processes the given audio input into a 3D binaural sound is the following:

aleft, aright hrtfmove2 aSignal, kAzimuth, kElevation, $HRTF_LEFT, $HRTF_RIGHT

where aleft and aright are the output signals for the left and for the right channels, aSignal represents the input stimulus, kAzimuth is the azimuth angle, kElevation corresponds to the elevation angle (set by default to 0 in our game), while $HRTF_LEFT and $HRTF_RIGHT represent the left, respectively the right HRTF spectral data files.

All the 3D sounds have been synthesized at 0 degrees elevation in the median plane, at the level of the listener's ears and have been delivered through stereophonic headphones [107], [108], [110].

The sound is generated in the CSound QT application. The command instr 1 generates the white noise. The variable "azimuth1" represents the value of the angle from the horizontal plane, while "kvalw1" is the proportion of white noise that will be perceived by the listener, in direct correspondence to the direction of the sound source in space. The variables "azimuth1" and "kvalw1" are calculated in the C# application and directly linked to the "whitepink.csd" file.

```
instr 1

kAzimuth chnget "azimuth1"

kElevation = 0   ;chnget "elevation"

kAmp chnget "ampli1"

kAmpVar = 0.3

kSinLFO    oscili kAmpVar, 1/2

kSquareLFO oscili kAmpVar, 1/4, 3

kValue chnget "kvalw1"

aSignal noise  kAmp * (1+kSinLFO*(kAmpVar+kSquareLFO)), 1

aleft, aright   hrtfmove2   aSignal, kAzimuth, kElevation, $FILE1, $FILE2 ;[,ioverlap, iradius, isr]

outs aleft*kValue, aright*kValue

endin
```

Similarly, the command instr 2 generates the pink noise. The variable "kvalp1" represents the proportion of pink noise corresponding to the angle "azimuth1" in the horizontal plane.

```
instr 2

kAzimuth chnget "azimuth1"

kElevation = 0   ;chnget "elevation"

kValue  chnget "kvalp1"
```

```
kAmp chnget "ampli1"

awhite unirand 2.0

white = awhite - 1.0

apink  pinkish kAmp*awhite, 1, 0, 0, 1

aleft, aright  hrtfmove2  apink, kAzimuth, kElevation, $FILE1, $FILE2 ;[,ioverlap, iradius, isr]

outs aleft*kValue, aright*kValue

endin
```

5.3. APPLICATIONS DEVELOPED FOR THE EXPERIMENTS

The applications described in this section have been developed in the Microsoft Visual Studio integrated development environment [121], using the C# programming language [122]. Also, they employ the auditory stimuli described in the previous subchapter.

5.3.1. The Binaural Navigation Test application

The Binaural Navigation Test application (Fig. 5.6.) has been designed with the purpose of conducting the pre-test and post-test sessions of the experiments described in this chapter.

In this application, the subjects were required to identify the position of a 3D sound source by freely navigating (using the touchpad or mouse movement interaction modality) from the starting position (which was randomly generated on the margin of a circle of 150 pixels radius) to the target sound source (situated exactly in the center of the circle). The audible area was a circle of 200 pixels radius around the source position.

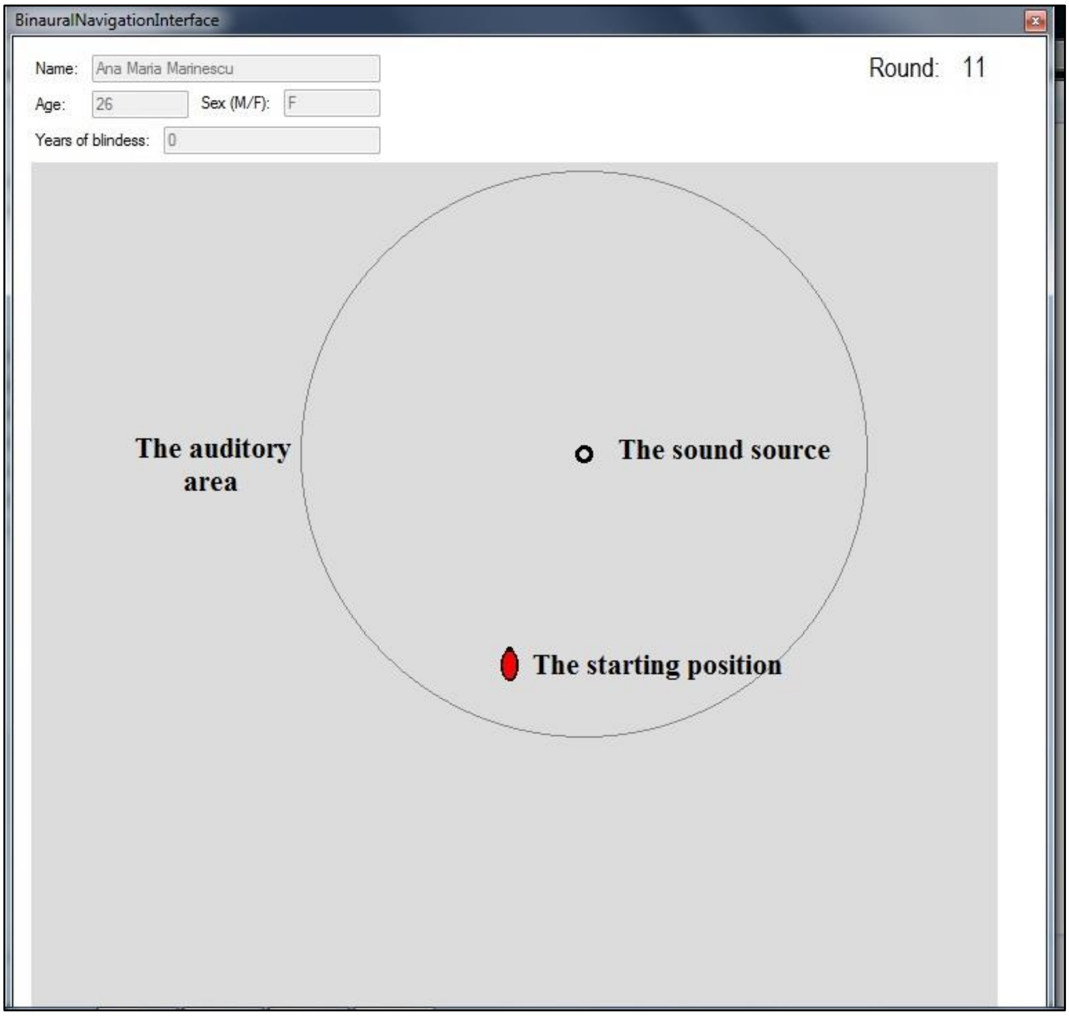

Fig. 5.6. The Binaural Navigation Test

The sonification technique for the Binaural Navigation Test application used the inversely proportional encoding of distance, so that the sound intensity increased as the user got nearer to the source and decreased as he got farther from it (reaching 0 outside the audible area of 200 pixels).

The presented level of the sounds was set to be comfortable for each listener, with maximum values around 65-70 dB SPL. As the user got nearer to the target source, the sound level increased and, as he or she got farther, the perceived amplitude decreased until complete silence (for the locations outside the 200 pixels auditory range). The formula that calculates the perceived volume of the target sound source is the following:

$$\text{Gain factor} = \begin{cases} 0, & d > dmax \\ GFMIN + (GFMAX - GFMIN) * \left(1 - \frac{d}{dmax}\right)^2, & 0 \le d \le dmax \end{cases} \tag{5.2}$$

Where d is the current distance between the position of the listener and the target sound source, *dmax*=150 pixels, *GFMIN*=0.05 (the minimum gain factor), *GFMAX*=1 (the maximum gain factor).

Sound intensity increases quadratically as the user approaches the target sound source. The dependency between sound intensity and distance is presented graphically in Fig. 5.7.

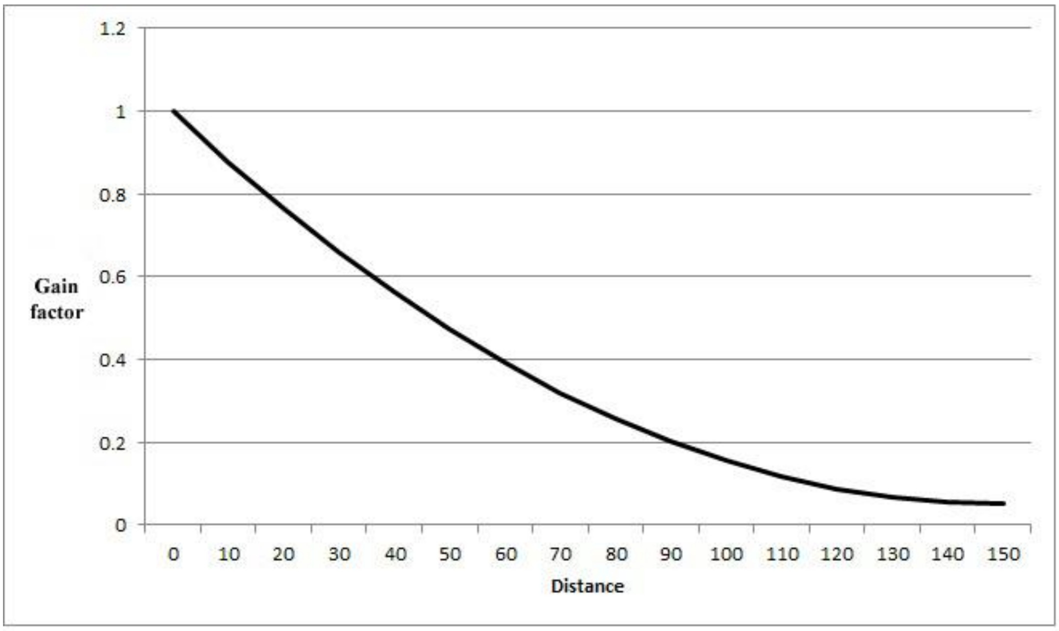

Fig. 5.7. The relationship between sound intensity and distance

Obviously, the formula presents a slight variation over the classically used inverse square low [123], being rather close to a linear decrease for our range of distances. We chose this encoding of distance because in our first prototypes and tests, this formula appeared to provide a better feeling of distance. While we didn't make any performance measurements to compare it with the inverse square law, all of the test users (who tested our prototypes during development; not the same as the subjects involved in the experiment) appreciated that this encoding allows for an easier evaluation and understanding of distance, in the range 0-150 pixels, at comfortable sound levels [109].

5.3.2. The Visual-Auditory Perceptual Training application

The Visual-Auditory Perceptual Training application [109] offers visual and auditory feedback about the perceived direction of the sound in space.

This application has been used in the experiment with sighted subjects. It has three modules (or constituent functionalities of the application):

a) A module where the subjects were required to freely move the mouse/touchpad cursor inside a circle and simultaneously hear the sound that corresponded to the angle between the center of the circle (considered as the listener’s fixed position) and the location pointed inside the circle, either under the white-pink noise synthesis or the “ding” sound listening conditions (Fig. 5.8.).

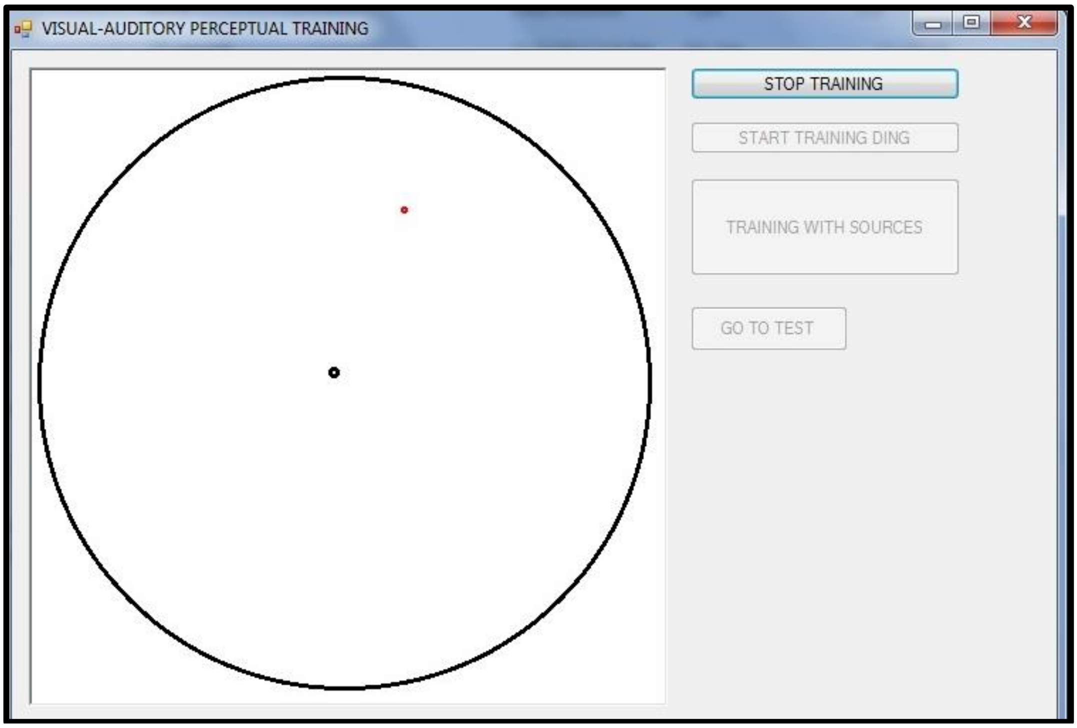

Fig. 5.8. The free listening module of the Visual-Auditory Perceptual Training application

b) A module that has 2 levels of difficulty and presents a sound discrimination procedure in which the number of virtual sound sources is restricted to four (for the first level) and eight (for the second level), for both types of sounds (Fig. 5.9.). In the first level, four virtual sound targets are randomly generated in each of the four quadrants of the auditory space. In the second round, the number of targets extends to eight (two sound sources in each quadrant). For each level, all the auditory targets appear on the screen as small circles of 5 pixels radius. The subjects are required to listen to the binaural sounds corresponding to the randomly generated 4 or 8 source targets and asked to indicate the perceived direction by clicking on the small circles assigned to the specific angular positions. When the listeners make a correct choice, the small circle corresponding to the current sound target disappears, reducing thus the searching range for the next targets.

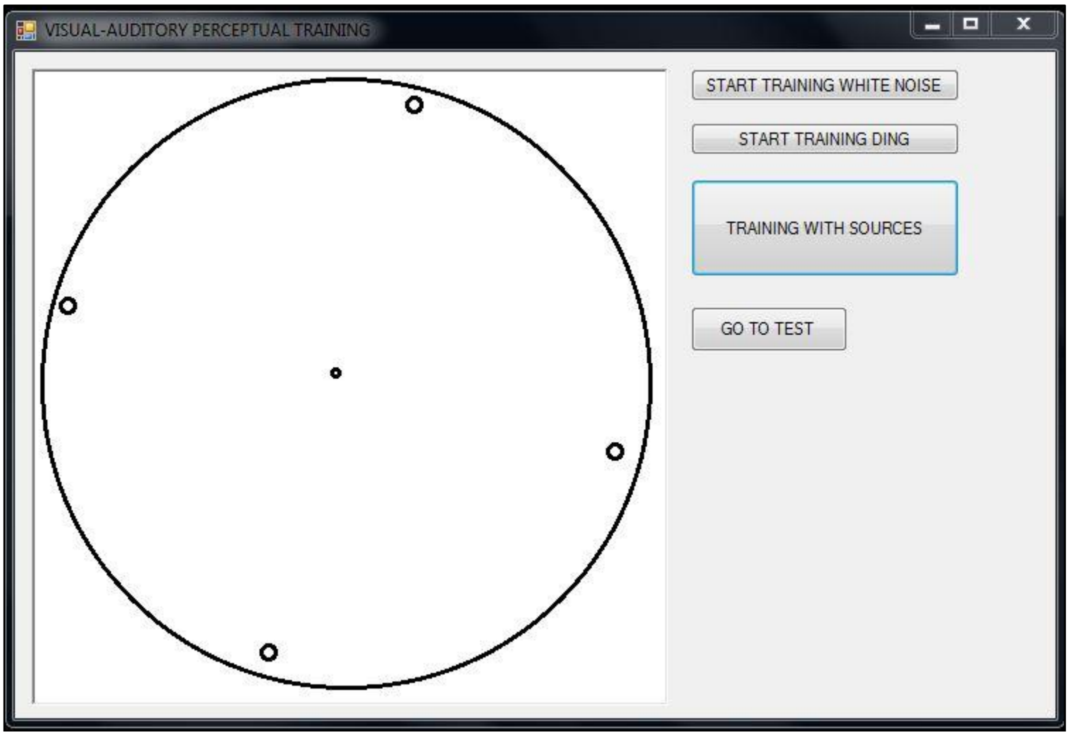

Fig. 5.9. The training application for the sighted subjects

c) A sound localization module that requires the listeners to point inside a circle the perceived direction of the incoming sounds that are presented via headphones. After the stimuli onset (after the sound is played), the users are asked to listen carefully to the target sounds in order to recognize their spectral characteristics, directionality and other features and consequently click inside the circle to indicate their apparent location. The subjects are offered immediate auditory and visual feedback, as the correct direction of the current sound stimulus is presented on the screen (colored in green), together with the subject' choice (colored in red) (Fig. 5.10.). An emulation of the target sound is simultaneously rendered over headphones in order to recalibrate the hearing and the visual senses and to create a solid association between the auditory perception of the 3D sounds and the correct direction of their source. Thus, the subject can see on the screen the correct direction of the sound source, as well as listen on the headphones to the 3D sound that corresponds to that particular direction.

The application has 10 consecutive rounds where the target auditory stimuli are the white-pink noise combination and another 10 rounds that employ the "ding" sound. The sound is rendered continuously, so that the subjects are offered a complete perception of the virtual auditory space.

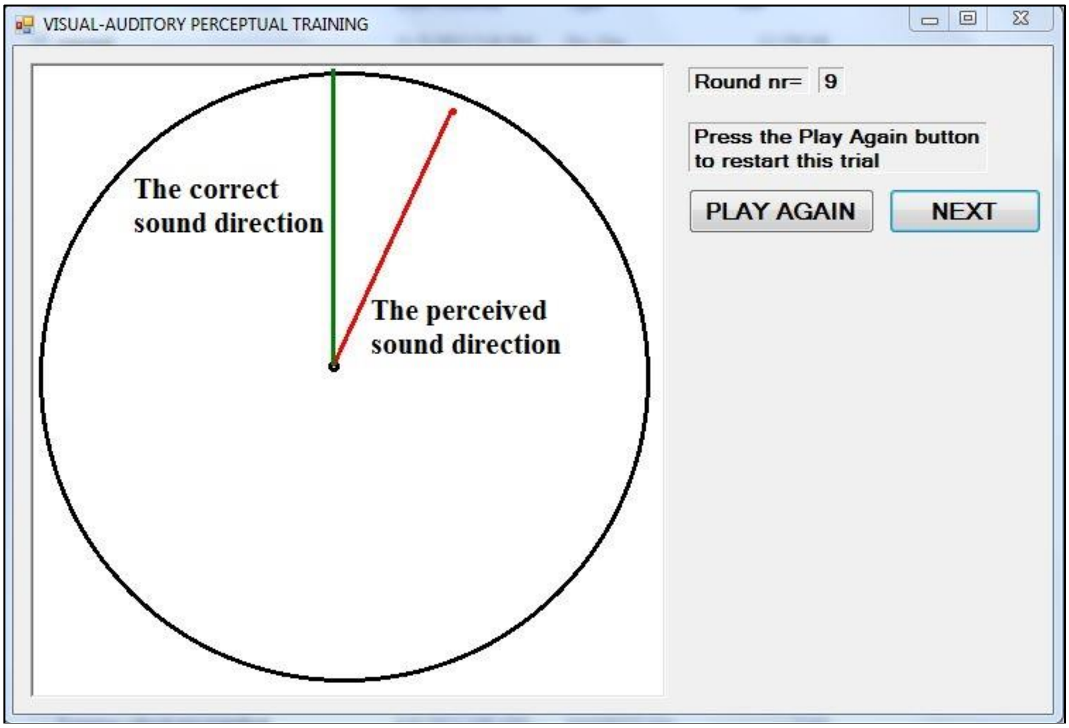

Fig. 5.10. Visual feedback concerning the correct direction of the sound source

5.3.3. The Haptic-Auditory Feedback Training application

The Haptic-Auditory Feedback Training application [107], [110] provides the listeners both auditory and haptic feedback about the perceived direction of the sound in space.

This application has been used in the experiment with visually impaired subjects. It requires the subjects to listen to different sound stimuli (corresponding to the directions 0, 30, 60, 90, 120, 150, 180, 210, 240, 270, 300 and 330 degrees) and to indicate the perceived sound direction using the conventional hour hand of the clock (for example, 12 o'clock for 0 degrees to the front, 3 o'clock for 90 degrees to the right or 6 o'clock for 180 degrees to the back) (Fig. 5.11.). The 3D sounds are presented continuously, in order to offer the listeners a complete perception of the auditory space. Consequently to the response, the subjects receive perceptual feedback about the correct direction of the sound source through a series of vibrations produced by the haptic belt they are required to wear on the head, paired with auditory feedback, in order to make an effective association between the current 3D sound and the angular direction it is originating from.

The C# command that transmits the vibration to the haptic belt is the following:

```
SerialPort port = new SerialPort("COM18", 9600, Parity.None, 8, StopBits.One);
String command;
if (index < 10) command = "S0" + index.ToString() + "400400\x0a";
else command = "S" + index.ToString() + "400400\x0a";
port.Write(command);
port.Close();
```

where index represents the number of the vibration motor from the haptic belt (corresponding to the direction of the 3D sound source in space) that is played with vibration type 40 for 4 times.

The subjects who have some degree of residual vision can also receive visual feedback - the correct direction of the sound source is presented graphically on the screen (colored in green), together with the listener's choice (colored in red) and auditory feedback.

The application has 12 rounds in which the listeners are required to listen to 3D binaural sounds that consist of the white-pink noise combination and to indicate the perceived direction of the sound, followed by another 12 rounds that employ the "ding" sound as the main test stimulus. As the head does not have a perfect spherical shape and as the head sizes and shapes vary among individuals, the experimenter needs to arrange the position of the actuators from the haptic belt on the heads of the participants at each round, making sure they match their corresponding angular locations. The vibration motors are numbered and the experimenter uses an arrangement scheme of the correct placement of the actuators around the subject's head, so that the motor number 1 should be placed straight in the front, number 7, in the back, 4, to the right and 10, to the left. The rest of the vibration motors are placed in-between the reference positions, ensuring that they are fixed at equal distances one from the other.

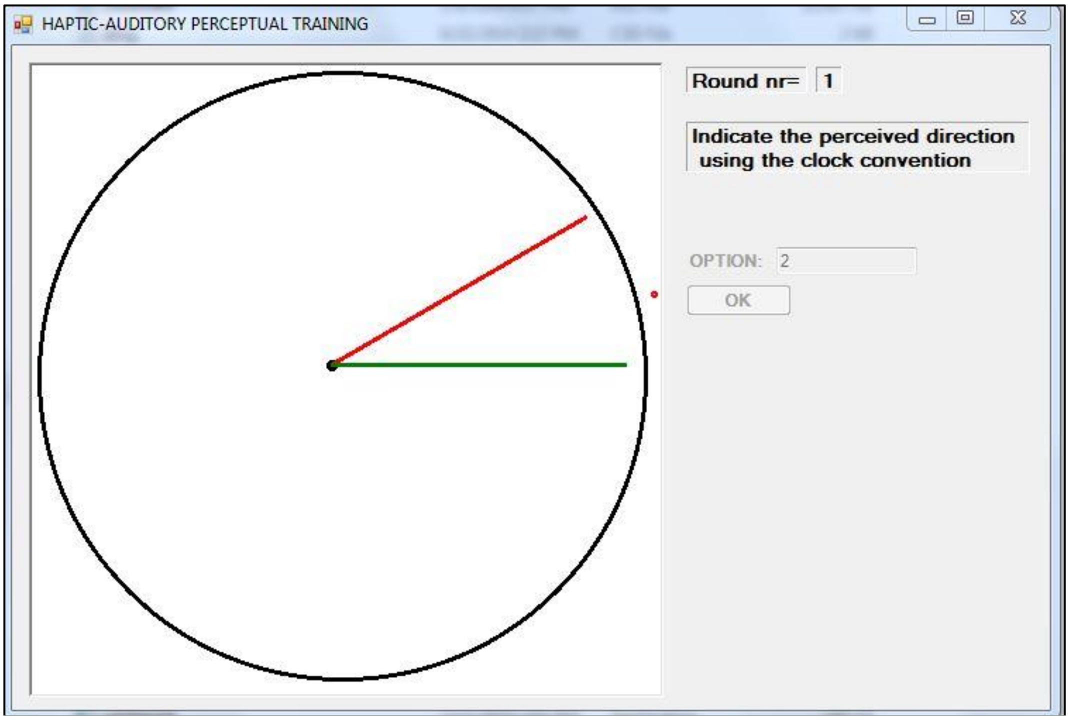

Fig. 5.11. The Haptic-Auditory Perceptual Training Application

5.4. THE SOUND LOCALIZATION EXPERIMENTS

This study comprised several sound localization tasks. For both groups of subjects (sighted and visually impaired), the experimental procedure included a pre-test session (where the spatial auditory resolution of the subjects has been evaluated using the Binaural Navigation Test application), a training session (aimed at helping the subjects to adapt to altered listening conditions, such as the use of 3D binaural sounds synthesized with non-individualized HRTFs) and a post-test session that is identical to the pre-test procedure, having as purpose to assess the level of sound localization improvement achieved as a result of training. The experiment with the sighted subjects took place in Ploieşti, Romania in December 2014, while the experiment with the visually impaired individuals has been conducted in Györ, Hungary, in February 2015 [110].

5.4.1. Target groups

Nine sighted subjects (6 women and 3 men living in Ploieşti, Romania) and another nine visually impaired individuals: 6 women and 3 men living in Györ, Hungary, with a percent of residual vision ranging from 0% to 20% - one subject was congenitally blind (0% residual vision), another was congenitally visually impaired (15% residual vision), while the other 7 were late-onset visually impaired, suffering from certain forms of visual impairments for 6 to 20 years.

All the subjects were naïve to the purpose of the experiment, as they have not listened to 3D binaural sounds nor did take part in any sound localization test before. Both groups reported normal hearing and the sighted subjects reported normal or corrected-to-normal vision.

5.4.2. The experimental procedure

THE PRE-TEST AND POST-TEST SESSIONS OF THE EXPERIMENTS

In the pre-test and post-test sessions of both experiments, the sound localization performance and navigational skills of the sighted and visually impaired subjects have been tested using the Binaural Navigation Test application (Fig. 5.12, Fig. 5.13) [107], [109], [110].

Both the pre-test and post-test sessions consisted of two blocks of trials. Each block of trials had 20 rounds - the rounds numbered from 1 to 5 and from 11 to 15 used the white-pink noise sound combination, while the rounds numbered from 6 to 10 and from 16 to 20 employed the “ding” signal as the main auditory stimulus.

Fig. 5.12. Sighted subject during the pre-test session

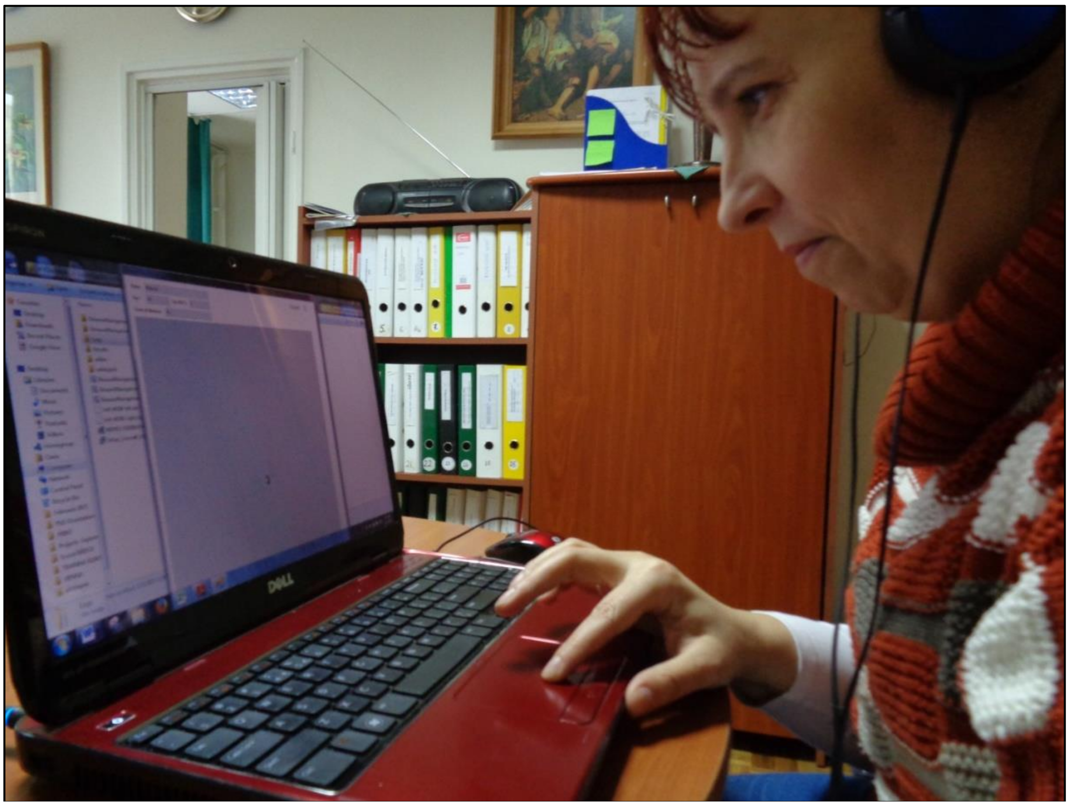

Fig. 5.13. Visually impaired subject during the pre-test session

The studied parameters were:

- **P1: The ratio of the distance travelled by the listener to the minimum possible distance of 150 pixels (the radius of the circle that encompasses the target sound source);**
- **P2: The percent of correct travel decisions**, defined as movements effectuated towards the sound source, minimizing the distance between the user's virtual location and the position of the current sound target;
- **P3: The round completion time** (in seconds), i.e. the amount of time that has passed from the moment the listener started to perceive the sound until he finally reached the target.

THE TRAINING SESSION

In the case of the experiment with sighted subjects [109], the training session has been extended over the course of 2 days. In each day, the listeners were required to use the Visual-Auditory Perceptual Training application as follows:

a). The free listening module that helps the subjects to get accustomed to the perception of 3D sounds has been employed by the subjects as long as they considered it was necessary (in average for 3 minutes).

b). The sound discrimination module with increasing degree of difficulty and the sound localization application that requires the listeners to point inside a circle the perceived direction of the incoming sounds has been used twice (Fig. 5.14.).

Fig. 5.14. Sighted subject during the training session

In the case of the visually impaired subjects [107], [108], [110], [111], the training session (which has been performed in 2 consecutive days, twice per day) has been composed of the following two tasks:

a) A free listening module, where the experimenter moved the mouse cursor inside a circle and the listener could hear the sound corresponding to the direction between the center of the circle (the virtual location of the listener) and the variable position of the mouse cursor, using as auditory stimuli both the white-pink noise combination and the "ding" sound.

b) The sound localization procedure implemented in the Haptic-Auditory Feedback Training application (Fig. 5.15.).

Fig. 5.15. Visually impaired subject during the training session

For the training session, the studied parameters were the reversal error rate (the percent of front-back and back-front confusions for each block of trials) and the mean angular precision error (defined as the unsigned difference between the correct direction of the sound source and the direction perceived by the listener).

The post-test session has been carried out using exactly the same stimuli and sound localization procedures as in the pre-test session.

5.5. THE BINAURAL NAVIGATION ANALYZER APPLICATION

The Binaural Navigation Analyzer application (Fig. 5.16.) [109], [110] allows real-time visualization and audio playback of the performance of the users for each round of the pre-test and post-test sessions (where the Binaural Navigation Test application has been used), for both experiments. Segments considered as "good movements" are colored in green, while those which correspond to "wrong movements" are colored in red. An alternative visualization paints the path travelled by the subjects in progressive grayscales, from source to destination. By pressing the "Browse log file" button, the evaluator can load the "log" files containing the results of the sound localization test sessions for each subject. The interface also displays the mean values for all the studied parameters, for the 10 rounds that used white/pink noise and for the 10 rounds that used the "ding" sound, as well as for all the 20 rounds. Finally, the "Export all data" button allows the data to be saved to an Excel file in order to be assessed for further analysis

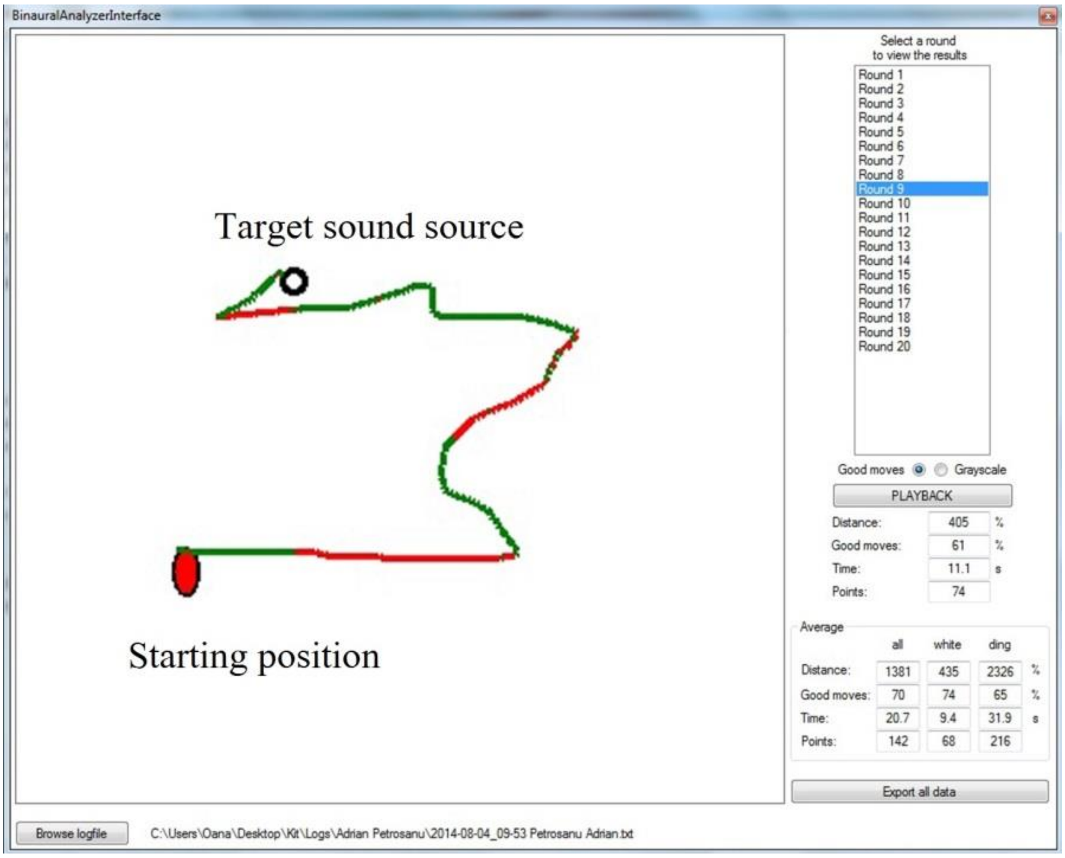

Fig. 5.16. The Binaural Navigation Analyzer

5.6. RESULTS

5.6.1. Statistical measures and tools used in the assessment of results

ANOVA is a statistical hypothesis test used to analyze the differences between groups in experimental data. In its simplest form, it tests whether the means of two or more groups are equal. The test result, obtained from the null hypothesis (that there is no difference between the studied groups and that they are part of the same population) and the sample, is statistically significant if it is unlikely to have occurred by chance. The results are statistically significant and the null hypothesis is rejected if a probability (p-value) is lower than a given threshold, known as the significance level [124].

The Student t-test is a statistical hypothesis that tests if two sets of data are different from each other. In our analysis, we used a "paired" or "repeated measures" t-test, which verifies if the difference between two responses (using a group of units that has been tested twice) has a mean value of zero [125].

The Wilcoxon signed-rank test is a statistical hypothesis that compares repeated measurements on a single sample, in order to study whether the mean ranks are different

[126]. In this case, the null hypothesis assumes that the medians of the two samples are identical.

5.6.2. Results of the pre-test and post-test sessions

For the rounds where the combination of white and pink noise in varying proportions according to the direction of the sound source in space has been used, the mean value of parameter P1 (ratio of the distance travelled by the listener to the minimum possible distance between the starting position and the target location) improved with 24.6% for the sighted group (although the results are not significant at $p<0.1$) and with 32% for the visually impaired group (in a Wilcoxon Signed Rank Test at $p<0.05$) (Table 5.2.). The results of the sighted subjects indicated a better evolution in the mean values of P1 (compared to their visually impaired counterparts), for both the pre-test and the post-test sessions of the experiment (for the pre-test session, the results are significant in an ANOVA test, $p=0.014$ and in a Student t-test where $t=2.74$, $p=0.007$. For the post-test session, the results are significant in the ANOVA test, $p=0.03$ and in a Student t-test where $t=2.37$ at $p=0.015$). Moreover, all the nine visually impaired subjects succeeded to obtain improvements in the mean value of P1 in the post-test session, whereas only 66% of the sighted participants recorded a higher level of performance for this parameter (Fig. 5.17.).

Table 5.2. Results of the pre-test and post-test sessions

Type of sound	Parameter	SIGHTED				VISUALLY IMPAIRED			
		Pre-test	Post-test	Level of improvement between the pre-test and the post-test session (%)	Percent of subjects who obtained an improvement in the post-test session (%)	Pre-test	Post-test	Level of improvement between the pre-test and the post-test session (%)	Percent of subjects who obtained an improvement in the post-test session (%)
White/ pink	P1	4.7	3.5	24.6	66	10	6.8	32	100
	P2 (%)	72.4	77.3	6.7	66	64	66.1	3.27	77
	P3 (seconds)	19.2	17.4	9.2	88	26.6	22	17.11	88
Ding	P1	7	3.9	43.6	100	12.9	7.5	41.61	100
	P2 (%)	70.6	74.6	5.5	77	61.2	65.1	6.33	100
	P3 (seconds)	24.8	16.9	31.8	100	32.8	23.5	28.21	88

In what concerns parameter P2 (the percent of correct travel decisions towards reducing the distance to the sound source), the mean value increased with 6.7% (from 72.4% to 77.3%) for the sighted group (the results are significant in an ANOVA test, $p=0.08$ and in a Student t-test for dependent means where $t=1.93$ at $p=0.08$) and with 3.2% (from 64% to 66.1%) in the case of the visually impaired subjects (the results are significant in a Student t-test where $t=2.46$ at $p=0.038$) (Figure 5.18.). The sighted individuals outperformed the visually impaired subjects in both the pre-test and the post-test sessions of the experiment (For the pre-test session, the results are significant in an ANOVA test, $p=0.002$ and in a Student t-test where $t= 3.55$, $p=0.0013$. For the post-test session, the results are significant in the ANOVA test at $p=8.9E-5$ and in a Student t-test where $t=5.18$ at $p=4.5E-5$). Similarly to P1, a higher percent of the visually impaired participants were able to improve their correct travel decision rate between the pre-test and the post-test sessions (77%, compared to 66% for the sighted group).

Regarding parameter P3 (the round completion time), its mean value decreased with 9.2% in the post-test phase (from 19.2 seconds to 17.4 seconds) in the case of the sighted subjects (although the results are not statistically significant at $p<0.1$) and with 17.1% (from 26.6 seconds to 22 seconds) for the visually impaired group (the results are not statistically significant at $p<0.1$). The sighted subjects outperformed their visually impaired counterparts in the pre-test session ($t=1.35$, $p=0.09$), although the results of both groups converged in the post-test stage (in the ANOVA test, the differences were not statistically significant at $p<0.1$). However, in both groups, 8 of 9 subjects (88%) recorded a noticeable improvement as a result of perceptual adaptation for parameter P3.

For the rounds that used the narrowband "ding" signal as the main auditory cue, the mean value of parameter P1 improved with 43.6% for the sighted group (the results are statistically significant in an ANOVA, $p= 0.05$ and in a Student t-test where $t=-2.83$ at $p=0.02$) and with 41.6% for the visually impaired group (the results are statistically significant in an ANOVA test, $p=0.04$ and in a Student t-test where $t=-4.44$, $p=0.001$). The results recorded by the sighted subjects showed a better evolution in what concerns the mean values of P1 (compared to the visually impaired individuals), for both the pre-test and the post-test sessions (for the pre-test session, the results are significant in an ANOVA test, $p=0.03$ and in a Student t-test where $t= 2.29$, $p=0.017$. For the post-test session, the results are significant in the ANOVA test, $p=0.01$ and in a Student t-test where $t=2.7$ at $p=0.007$). At the same time, all the sighted and the visually impaired subjects (100%) succeeded to reduce the ratio of the distance travelled to the minimum possible distance between the starting position and the target sound source during the post-test trials.

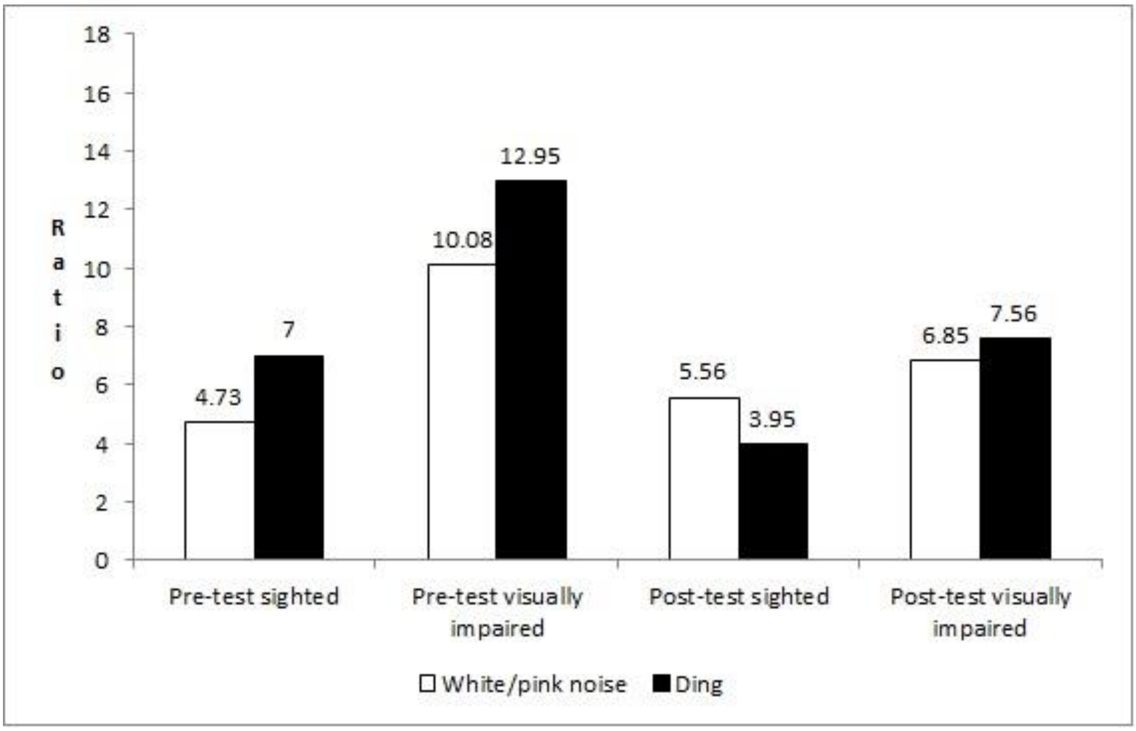

Fig. 5.17. Evolution of parameter P1 in the pre-test and post-test sessions, for the sighted and the visually impaired groups

Regarding parameter P2, the mean value increased with 5.5% (from 70.6% to 74.6%) for the sighted group (the results are significant in an ANOVA test, $p=0.09$ and in a Student t-test where t=2.71 at p=0.026) and with 6.3% (from 61.2% to 65.1%) in the case of the visually impaired subjects (the results are statistically significant in a Wilcoxon Signed Rank Test at $p<0.05$) (Fig. 5.18). The sighted individuals' percent of correct travel decisions was higher than that of the visually impaired subjects in both the pre-test and the post-test sessions (for the pre-test session, the results are significant in an ANOVA test, p=0.0001 and in a Student t-test where t=4.4, p=0.00002. For the post-test session, the results are significant in the ANOVA test, p=0.002 and in a Student t-test where t=3.67 at p=0.001). Similarly to the rounds where the white/pink noise combination has been used, a higher percent of the visually impaired subjects improved the mean rate of parameter P2 after the training procedure (100% for the visually impaired group and 77% for the sighted group).

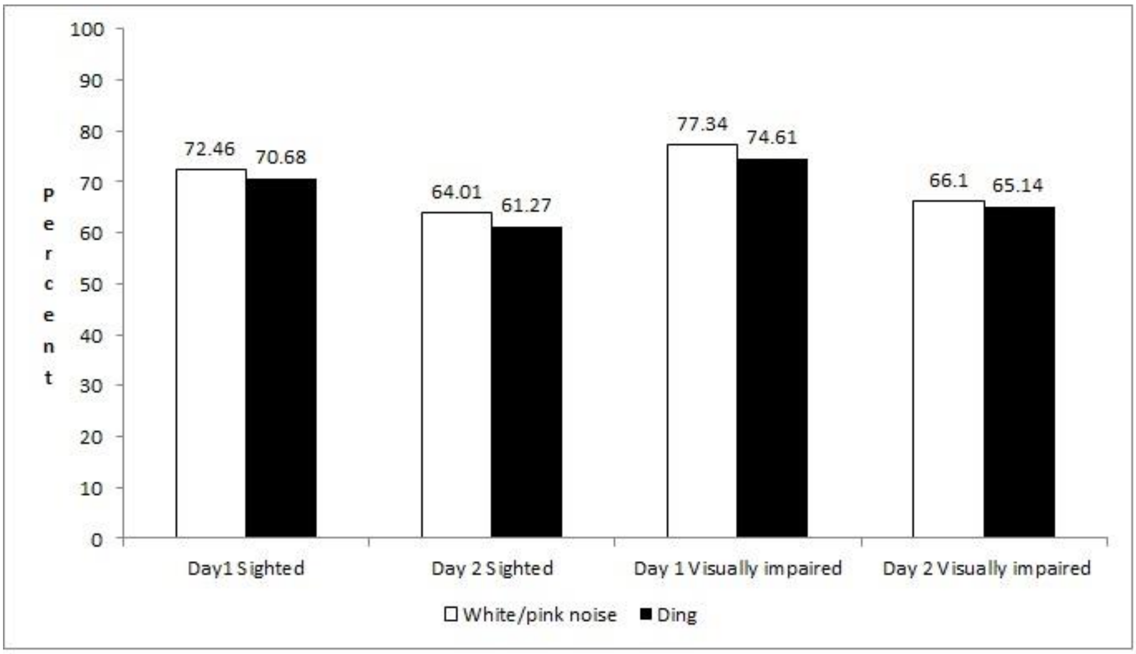

Fig. 5.18. Evolution of parameter P2 in the pre-test and post-test sessions, for the sighted and the visually impaired groups

In what concerns parameter P3, the decrease in the round completion time was of 31.8% (from 24.8 seconds to 16.9 seconds) for the sighted group (the results are statistically significant in an ANOVA test, p=0.02 and in a Student t-test for dependent means where t=-3.37 at p=0.004) and of 28.2% (from 32.8 seconds to 23.5 seconds) for the visually impaired group (t=-2.1 at p=0.03). In this case, the sighted subjects achieved better results for the mean round completion time than the visually impaired subjects in the pre-test session, although the results of both groups were not statistically different. Nonetheless, all the sighted subjects succeeded to complete the rounds quicker in the post-test trials, while only 8 of the 9 visually impaired users (88%) obtained a lower mean value of parameter P3 between the two test sessions (Table 5.3.) [107], [109], [110].

Table 5.3. Statistical comparison between the sighted and the visually impaired subjects in the pre-test and post-test sessions of the experiment

	Parameter P1		**Parameter P2**		**Parameter P3**	
Comparison between the sighted and the visually impaired groups	**White-pink noise**	**“Ding” sound**	**White-pink noise**	**“Ding” sound**	**White-pink noise**	**“Ding” sound**
Pre-test session	**ANOVA** p=0.014 **Student t-test** t=2.74 p=0.007	**ANOVA** p=0.03 **Student t-test** t=2.29 p=0.017	**ANOVA** p=0.002 Student t-test t=3.55 p=0.0013	**ANOVA** p=0.0001 **Student t-test** t=4.4 p=0.0002	**Student t-test** t=1.35 p=0.09	No significant difference between the groups
Post-test session	**ANOVA** p=0.03 **Student t-test** t=2.37 p=0.015	**ANOVA** p=0.01 **Student t-test** t=2.7 p=0.007	**ANOVA** p=8.9E-5 **Student t-test** t=5.18 p=4.5E-5	**ANOVA** p=0.002 **Student t-test** t=3.67 p=0.001	No significant difference between the groups	No significant difference between the groups

5.6.3. Results of the training session

For the rounds where the combination of white and pink noise in varying proportions according to the direction of sound has been employed, we recorded a reduction of 14.7% in the sound localization error (from 24.1 to 20.5 degrees) in the case of the sighted subjects and of 25.7% (from 37.3 to 27.7 degrees) (Fig. 5.19) in the case of the visually impaired individuals (t=-2.57 at p=0.01) (Table 5.4.). Even though the localization error was larger for the visually impaired group, their improvement rate significantly surpassed that of the sighted counterparts with more than 10%. The same situation is encountered when assessing the front-back confusion rate, as the sighted individuals obtained an improvement rate of 33.3% (from 8.3% to 5.5%), while the visually impaired subjects reduced the percent of reversal errors with 50% (from 12% to 6% - the results are statistically significant in an ANOVA test, p=0.02 and in a Student t-test where t=-3.71 at p=0.002) (Fig.5.20). We can notice also that the number of individuals who achieved a higher sound localization accuracy is larger in the visually impaired group (with 11% more in the case of precision localization errors and 33% in the case of reversal misjudgments). The visually impaired subjects reported higher angular precision errors than the sighted participants in both days of training (for the first day of training, the differences are statistically significant in an ANOVA test, p=0.01 and in a Student t-test where t=2.69 at p=0.008). On the other hand, the ANOVA test revealed that the front-back localization judgments are comparable for the sighted and the visually impaired groups, for both days of training (the differences are not significant at p<0.1).

Table 5.4. Results of the training session

Type of sound	Parameter	SIGHTED				VISUALLY IMPAIRED			
		Day 1	Day 2	Level of improvement between first and the second day of training (%)	Percent of subjects who obtained an improvement in the second day of training (%)	Day 1	Day 2	Level of improvement between first and the second day of training (%)	Percent of subjects who obtained an improvement in the second day of training (%)
White/pink	Angular precision error (degrees)	24.1	20.5	14.73	66	37.3	27.7	25.7	77
	Reversal error rate (%)	8.3	5.5	33.37	77	12	6	50	100
Ding	Angular precision error (degrees)	44.7	45.8	-2.47	44	44.8	42.4	5.4	55
	Reversal error rate (%)	22.7	19.4	14.69	77	14.3	12.4	12.9	55

For the rounds that used the "ding" sound as the main auditory stimuli, the sighted subjects recorded an increase in the angular precision error of 2.4% (from 44.7 degrees to 45.8 degrees – the results are not statistically significant), while the visually impaired succeeded to obtain a decrease of 5.4% (from 44.8 degrees to 42.4 degrees). At the same time, the percent of subjects who achieved better angular localization accuracy is slightly higher for the visually impaired group (55%, compared to 44% in the case of the sighted group). In what concerns the front-back confusion rate, the level of improvement is higher for the sighted subjects (14.9%, from 22.7% to 19.4%). Also, the percent of individuals who were able to enhance their front-back auditory localization judgment was larger in the case of the sighted group (77%, compared to 55% for the visually impaired). The visually impaired subjects achieved a lower rate of front-back confusion errors than the sighted participants in both days of training

(In the first day, the results are significant in an ANOVA test, $p=0.09$ and in a Student t-test where $t=1.78$, $p=0.04$, while in the second day of training the differences between the groups are not statistically significant). Nonetheless, the angular precision performance is comparable for the sighted and the visually impaired groups, in both days of the training session (the differences are not significant at $p<0.1$) (Table 5.5.) [107], [109], [110].

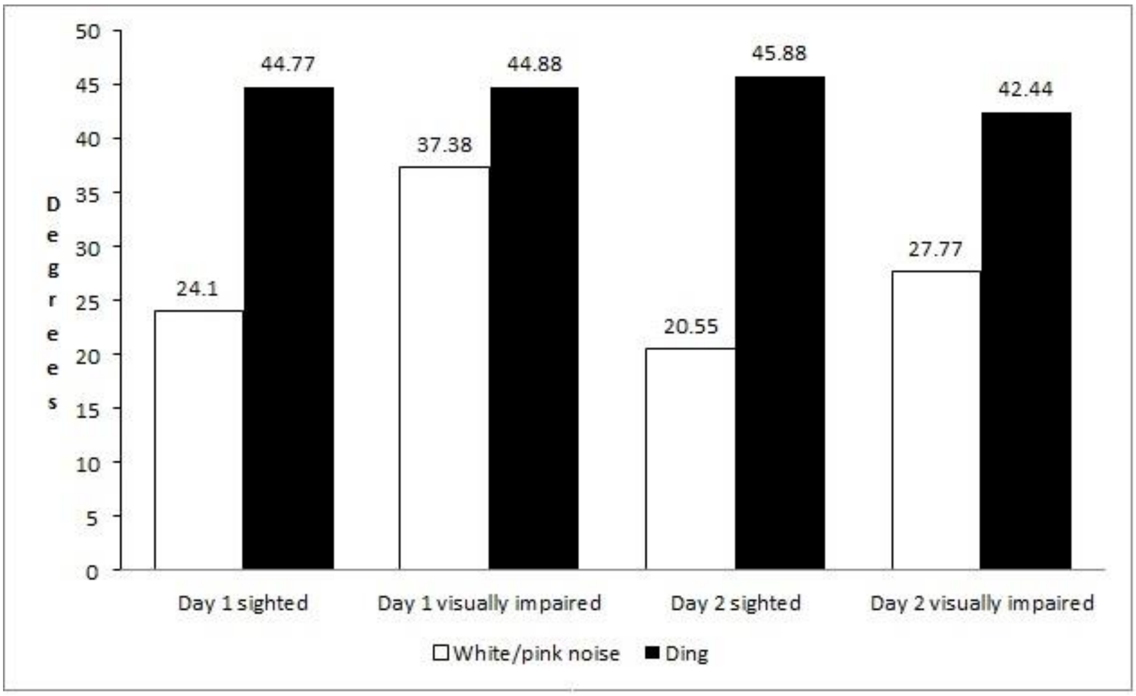

Fig. 5.19. Evolution of angular precision errors in both days of training, for the sighted and the visually impaired group

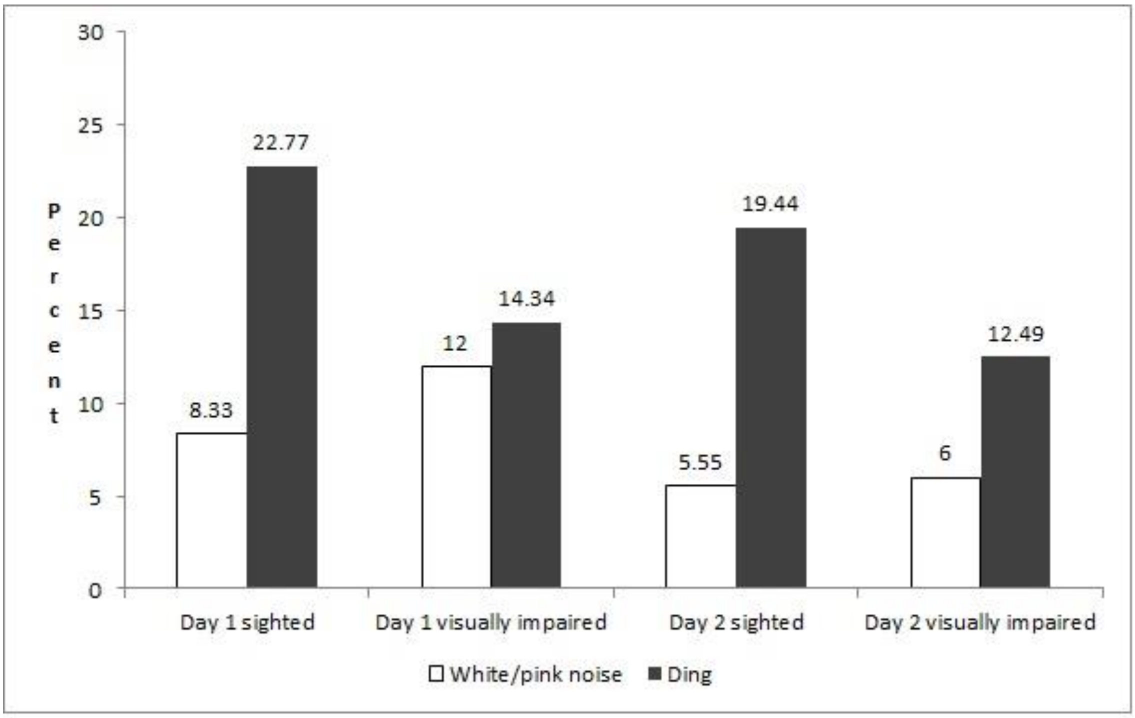

Fig. 5.20. Evolution of reversal error rates in both days of training, for the sighted and the visually impaired group

Table 5.5. Statistical comparison between the sighted and the visually impaired subjects in the two days of training

	Angular precision error		**Front-back confusion rate**	
Comparison between the sighted and the visually impaired groups	**White-pink noise**	**"Ding" sound**	**White-pink noise**	**"Ding" sound**
Day 1 of training	**ANOVA** p=0.01 **Student t-test** t=2.69 p=0.08 Better performance in the case of the sighted subjects	No significant difference between the groups	No significant difference between the groups	**ANOVA** p=0.09 **Student t-test** t=1.78 p=0.04 Better performance in the case of the visually impaired subjects
Day 2 of training	No significant difference between the groups	No significant difference between the groups	No significant difference between the groups	No significant difference between the groups

5.6.4 Quadrants analysis

This subsection presents the analysis of the sound localization accuracy of both groups of subjects in each of the 4 quadrants of the auditory space (Fig. 5.21).

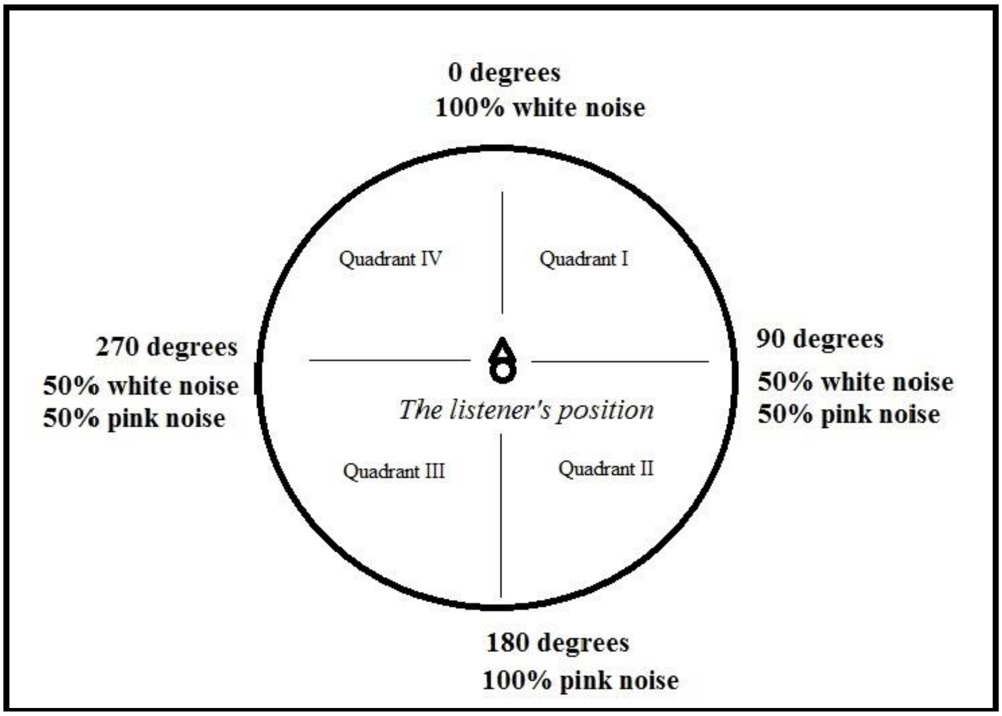

Fig. 5.21. The azimuthal range for each quadrant of the auditory space

QUADRANTS ANALYSIS FOR THE EXPERIMENT WITH SIGHTED SUBJECTS

The highest rate of front-back confusions (60% of the total number of reversal errors) has been obtained in the rear hemifield (quadrant 2 and 3) when listening to the mix of white-pink noise and in the frontal hemifield (56%) in the case of the "ding" sound. Moreover, for both types of sound stimuli, the largest mean localization errors have been recorded in the fourth quadrant (32° for the white-pink noise combination and 60° for the "ding" sound). Also, the best localization performance was recorded in the second and in the third quadrants (in the rear hemifield) – mean localization errors of 19.5° [109].

QUADRANTS ANALYSIS FOR THE EXPERIMENT WITH VISUALLY IMPAIRED SUBJECTS

The lowest front-back confusion rate has been achieved for the sound targets situated in the second and in the fourth quadrants (Table 5.6.), while the highest percent of front-back confusion errors has been recorded for the sound sources located in the first quadrant (when listening to the white-pink noise sound stimuli) and in the third quadrant (when using the "ding" sound as the main auditory cue). In what concerns the precision error, the highest errors have been recorded in the first quadrant (for the directions 0, 30 and 60 degrees, for both types of stimuli), while the best localization performance has been obtained in the third quadrant (for the white-pink noise stimuli) and in the second quadrant (for the "ding" sound).

For the white/pink noise combination, the largest rate of front-back confusions has been obtained at 0 degrees to the front (15%), while the lowest percent of reversal errors (0%) occurred at 150 degrees (in the second quadrant) and at 210 degrees (in the third quadrant). The lowest angular localization accuracy has been obtained at 0 degrees to the front, while the best performance was recorded at 180 degrees in the back (15 degrees).

For the “ding” sound, the largest rate of front-back confusions has been recorded also at 0 degrees to the front and at 180 degrees in the rear (16%), while the lowest percent of reversal errors (2%) occurred at 60 degrees (in the first quadrant). The lowest angular localization accuracy has been obtained at 30 degrees to the front, while the best performance (angular error of 30 degrees) was recorded at 60 degrees (in the first quadrant) and at 300 degrees (in the fourth quadrant) [107], [110].

Table 5.6. Distribution of front-back confusions and angular localization errors in the four quadrants of the auditory space

QUADRANT	Front-back confusion rates White/Pink noise (%)	Angular precision errors White/Pink noise (degrees)	Front-back confusion rates “Ding” sound (%)	Angular precision errors “Ding” sound (degrees)
Quadrant I	18.7	55	10	45.6
Quadrant II	2	28.6	6.9	37.3
Quadrant III	9.3	25.3	13.1	41.3
Quadrant IV	3.1	32	3.1	40.6

5.6.5. Results of the usability study

After the post-test session, the subjects were required to complete a usability questionnaire in which they expressed their opinions about the sound localization experiment they took part in. We hereby present the content of the questionnaire:

1. **How often do you use the headphones for the computer or for the mobile phone?**
a. Very often b. Often c. Neither often, nor rarely d. Rarely e. Very rarely

2. **I consider that the sonification technique based on the combination of white and pink noise helped me to avoid front-back confusions and to identify the sound sources from the front and from the back.**
a. Very much b. Much c. Neither much, nor little d. Little e. Very little

3. **I consider that the training session helped me to improve my sound localization skills.**
a. Very much b. Much c. Neither much, nor little d. Little e. Very little

4. Answer with TRUE or FALSE.

CLASSIFICATION	SENTENCE	TRUE	FALSE
MOTIVATION AND INTERACTION	The sound localization application is motivating and interesting		
	I consider that this application would be interesting for the development of an assistive device for the visually impaired people		
	I would like to use the software for training my sound localization abilities		
USING THE APPLICATION	The application's interface is easy to be learnt and used		
	The application is intuitive and I have learned to use it quite quickly.		
3D SOUND PERCEPTION	The 3D sounds are efficient for localizing the sound source in space		
	It was easy to identify the sound source using the white-pink noise combination		
	It was easy to identify the sound source using the "ding" sound		
	It was easy to identify the sources from the front and from the back using the white-pink noise combination		
	It was easy to identify the sources from the front and from the back using the ding sound		

OPINIONS OF THE SIGHTED SUBJECTS

77% of the subjects considered that the method based on the combination of white and pink noise in varying proportions that is aimed at reducing the incidence of front-back confusions is highly efficient, while 22% of them believed that this is a quite good technique, as it enables them to discriminate between the sounds coming from the back or from the front. Also, the participants appreciated that the training session helped them to get accustomed with the perception of 3D sounds and to improve their spatial auditory performance (55% of the subjects believed that the training session was very useful, while 44% admitted that it was a good approach for adapting to altered auditory cues). The subjects unanimously appreciated that the Binaural Navigation Test application is challenging, motivating and interesting and that it would serve for the future development of an assistive device for the visually impaired people. Furthermore, they considered that the software is interactive and that its interface is easy to be learned and used. Moreover, all the subjects agreed that the 3D sounds are efficient for sound localization in a virtual auditory environment and that the white-pink noise synthesis provides a better spatial resolution than the "ding" sound. Consequently, they admitted that the front-back auditory discrimination has been facilitated by the spectral characteristics of the white-pink noise and subsequently that it was made more difficult when the "ding" sound has been employed [109], [110].

OPINIONS OF THE VISUALLY IMPAIRED SUBJECTS

77% of the subjects considered that the sonification technique based on the combination of white and pink noise helped them to avoid front-back confusions and to identify the sound sources originating from the frontal and from the rear plane. Also, 77% of the visually impaired participants expressed their satisfaction concerning the efficiency of the training session. Thus, they considered that the haptic-auditory perceptual feedback based training helped them to improve their sound localization performance and front-back disambiguation abilities. The subjects unanimously appreciated that the sound localization application is motivating and challenging, that it is easy to be learnt and used and that it can be the support for the future development of an assistive device for the visually impaired people. 88% of the subjects considered that the 3D sounds are efficient for localizing the sound source in space and that it was easy to identify the sound source by using the white-pink noise combination. In addition to this, 66% of the users appreciated that it was easy to identify the sound sources using the “ding” sound and that the combination of white and pink noise in varying proportions enabled them to correctly identify the sound sources situated inside the cone of confusion. Nonetheless, only 44% of the subjects considered that the “ding” sound is efficient for avoiding the front-back confusion errors (Fig. 5.22.).

In a short interview that took place after the experiment, the subjects expressed their appreciation concerning the proposed application. They also argued that this can be the baseline for developing navigational audio games that would improve their sound localization abilities and that additional training would be highly useful for supporting enhanced orientation and mobility skills. Furthermore, one subject considered that the haptic belt can be used as a standalone device for improving navigation outdoors, by helping the blind people to keep a straight line when walking or by providing directional information [110].

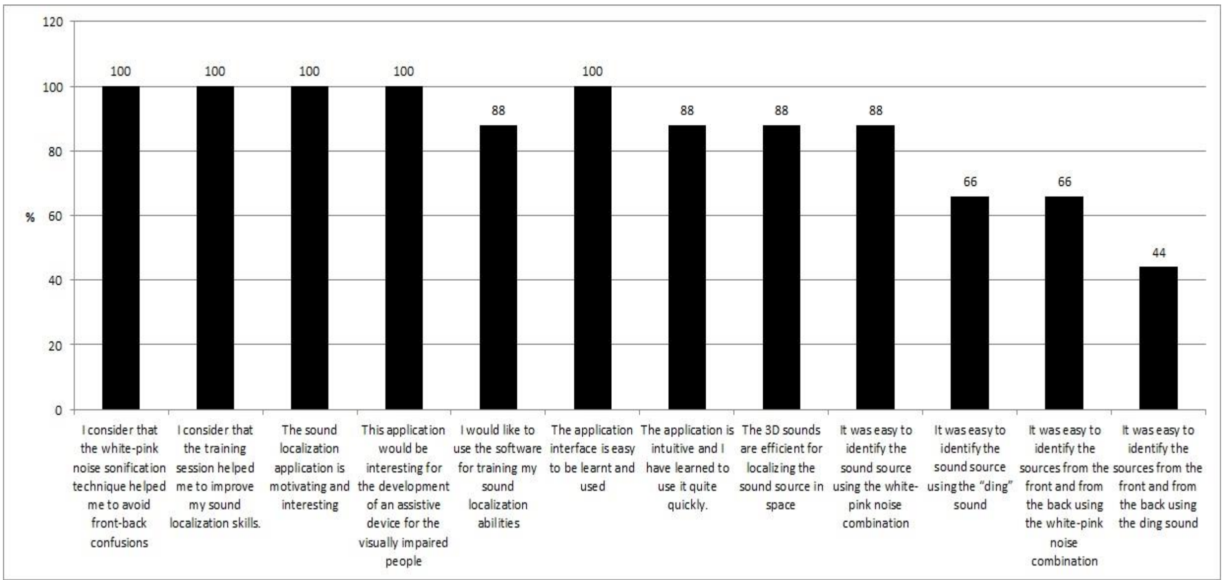

Fig. 5.22. Results of the usability study for the visually impaired group

5.7. DISCUSSION

The results provided by our experiments demonstrated a rapid improvement in the sound localization accuracy and front-back disambiguation for both the sighted and the visually impaired subjects. We consider that this is due to learning the procedure - how to identify the spectral characteristics of the sound (especially for the white-pink noise combination) and to the perceptual feedback based training method.

5.7.1. Discussion on the results obtained in the post-test sessions

In the post-test sessions of the experiments, the level of improvement recorded by the visually impaired subjects is higher than that of the sighted subjects (for parameters P1 and P3), for the rounds where the combination of white and pink noise in varying proportions has been used as the main auditory cue. For the rounds where the narrowband "ding" signal has been used, the degree of improvement is higher for the sighted subjects (for parameters P1 and P3), although the visually impaired participants succeeded to obtain a higher percent of improvement for P2 (the rate of correct travel decisions). For both types of sound stimuli, for all the three studied parameters (except for the "ding" sound, for parameter P3), the percent of visually impaired subjects who recorded a significant improvement in the post-test session of the experiment is equal or higher than that of the sighted individuals, demonstrating that the level of spatial auditory adaptation is larger in the case of the subjects who suffer from a certain degree of visual disability [109], [110].

5.7.2. Discussion on the results obtained in the training sessions

In the second day of training, the visually impaired subjects outperformed their sighted counterparts in the percent of angular precision and reversal error rate improvement. Also, the number of subjects who obtained a better sound localization performance as a result of training was higher in the visually impaired group (except for the rounds where the "ding" sound has been used, where the sighted subjects recorded a higher level of improvement than the visually impaired for the reversal error rate parameter).

We believe that the adaptation process that took place during the training session was concentrated on learning how to focus on the spectral characteristics of the auditory stimuli, as an indicator of the spatial position of the target source. Our subjects succeeded to develop a better process to map the auditory stimuli to the correct answer on a rapid time scale, fact that can be explained by an increased selective attention to the spectral profile of the sounds during the training phase. These results can be explained by the fact that the white and pink noises are more externalized than the narrowband "ding" sound, as the broadband noises contain much more spectral information that facilitates the localization process. Although they are not very natural sounds, the white and pink noises surpass the narrowband signals due to their enhanced directionality and laterality (externalization or out-of-the head perception). Also, the white and pink noises were continuous, offering a complete auditory perception, while the "ding" stimuli were discrete and repetitive [109], [110].

5.7.3. Comparison between the results obtained by both groups of subjects

The sighted subjects obtained a higher angular precision accuracy than the visually impaired individuals for the rounds where the white and pink noises have been used as auditory cues. However, both groups recorded similar reversal error rates in both days of the training session. In the case of the rounds where the narrowband "ding" signal has been employed, the angular precision error rates were comparable for both groups, although the visually impaired subjects made less reversal errors than the sighted participants.

For this type of stimuli, both the sighted and the visually impaired subjects who participated in our experiments obtained comparable reversal error rates, reaching a mean reversal error rate of 5.55% (the sighted), respectively 6% (the visually impaired) in the second day of training. These results demonstrate the efficiency of our method, proving that the combined spectral features of both types of noise enabled the listeners to differentiate between the sources located in frontal or rear position.

For the trials where the sonification technique was based on the use of the narrowband "ding" stimulus, the sound localization performance of both groups of subjects was weaker than for the rounds which employed the white/pink noise combination. The angular precision error rate was comparable for both groups of subjects. Nonetheless, the visually impaired listeners were able to better disambiguate the sources located on the cone of confusion, reaching a mean reversal error rate of 12.9% in the second day of training. This result demonstrates that the visually impaired individuals are able to identify the hemifield the sound source is originating from with higher accuracy than the sighted people, even when listening to narrowband auditory stimuli.

Nonetheless, a plausible explanation for the slightly poorer spatial auditory localization accuracy of the visually impaired subjects in the training and post-test session (compared to that of the sighted individuals) is the limited resolution of the haptic belt used for training (the vibration motors were placed at 30 degrees difference around the head), while the visual feedback (with perfect resolution) used by the sighted individuals provided a complete spatial perception of the environment [109], [110].

5.7.4. Comparison between this study and other similar experiments

The results of our experiment are better than those obtained by Blum et al [38], who recorded for their test group a mean precision error of 29 degrees and a rate of front-back confusions of 25% after the training session. Similarly, our results are comparable with those presented by Majdak et al [60], who recorded a precision error of 23.3 degrees before training and of 19.8 degrees after the visual-auditory feedback based adaptation procedure.

However, the almost identical localization accuracy improvements (3.6 degrees through our solution compared to 3.5 degrees in [60]) were achieved in our case in a training period more than 100 times shorter: an average of 40 minute of training (40 trials split in 2 consecutive days) in our case, versus an estimated average of 4500 minutes (600 to 2200 trials over an average of 20 days) as in [60].

In what concerns the front-back confusion rate, our results are better than those obtained by Parseihian and Katz [42], who recorded a reduction in the front-back confusion rate from 25-27% to 11% in the post-test session of the experiment. In their experiment, the12 minutes adaptation task (which has been performed for three days) consisted in a game where the subjects were required to search for animal sounds hidden around them, using a hand-held position-tracked ball. Also, we recorded a lower incidence of reversal errors (for both groups of subjects) than Zahorik et al [55], who obtained a reduction from 38% to 23% in the front-back confusion rate after 2 training sessions of 30 minutes in which the listeners were provided auditory, visual and proprioceptive feedback. Moreover, our results are better than those of Wenzel [127] who recorded a mean front-back confusion rate of 32% (ranging from 20% to 43%) in the virtual auditory environment and comparable with her results under free-field listening conditions (mean reversal error rate of 6.5%, ranging from 2% to 10%). In addition, the performance of our subjects is higher than that of the subjects who participated in Padersen and Jorgensen's experiment [128] who recorded a front-back confusion rate of 21.3% for 250 ms long virtual white noise stimuli. Also, our results are even better than the reversal error rate obtained by them under free-field listening conditions (9.1%), for the same type of auditory cues [109], [110].

5.8. CONCLUSIONS

This study demonstrated that the human auditory system is able to quickly adapt to altered hearing conditions, such as listening to 3D binaural sounds filtered with non-individualized HRTFs. The adaptation process was the result of a multimodal perceptual association between the hearing and the visual (in what concerns the sighted subjects) or tactile sense (being the case of the visually impaired participants). Both the sighted and the visually impaired listeners succeeded in improving their sound localization performance by reducing the reversal and precision error rates for both types of stimuli (with significant better results for the rounds that employed the synthesis of white and pink noise in varying proportions, according to the direction of the sound source in space). Although the sighted subjects outperformed their visually impaired counterparts in most of the required tasks, the level of improvement of the visually impaired participants in the second day of training, respectively in the post-test session of the experiment was generally higher than that of the sighted individuals. Also, the percent of subjects who recorded a significant improvement as a result of training is higher in the visually impaired group. Thus, the untrained sighted subjects recorded higher results than the visually impaired users in the post-test session, for all the three studied parameters, when listening to the combination of white and pink noise in varying proportions. Also, for the "ding" sound, the untrained sighted were more proficient than their counterparts, for parameters P1 and P2. Only for P3, the trained blind subjects recorded a slight advantage (they obtained a mean completion time of 23.5 seconds, whereas the sighted users had a mean rate of 24.8 seconds).

The results of our test and training sessions prove that the sighted and visually impaired subjects have been able to use 3D binaural sounds synthesized from non-individualized HRTFs as the only means for navigating in a virtual auditory environment. Besides, the results of the post-test session demonstrate that the subjects enhanced their orientation and mobility skills, improved their directional decision-making abilities and recalibrated the spatial resolution when navigating in a virtual auditory environment, for both types of sounds [109], [110].

CHAPTER 6
AUDIO GAMES FOR TRAINING SOUND LOCALIZATION, EXPERIMENTS AND RESULTS

This chapter presents the design of two audio games and the results of two experimental procedures with both sighted and visually impaired subjects. The details of implementation, the experimental methodology and the results have been published in [59], [129] and [130].

6.1. EXPERIMENTS WITH SIGHTED SUBJECTS

6.1.1. Audio games developed for the experiments

The audio games described in this section have been developed in the Visual Studio integrated development environment, using the C# programming language. The goal of the games consists in exploring a virtual auditory environment while listening to 3D binaural sounds that guide the player towards finding 5 hidden objects as quick as possible. The games end when the 5th object has been discovered.

The virtual environment is designed as a square-shaped graphical field in which the players are required to freely navigate by moving the mouse cursor in order to discover the 5 invisible 3D sound targets. By pressing the left button of the mouse, the players can stop from navigating. By rotating the scroll wheel of the mouse, they can simulate a shift in the position of their head towards the sound source. For example, if the sound source is situated at 90 degrees to the left, by moving the head in that direction, the players can perceive the signal as coming from the front. To restart navigation, the subjects are required to press the right button of the mouse.

For the first game (Fig. 6.1.), the starting point is the center of the 500x500 pixels virtual field. The position of the first sound target is randomly selected within a range of 200 pixels around the starting point. The next 4 targets are arbitrarily positioned in the same 200 pixels range in respect with the location of the previous object. Only one sound source can be perceived at a time. As the game advances and the subjects discover the targets, they became visible on the screen and the next source starts to emit sound. The game ends when the locations of all the 5 targets have been identified.

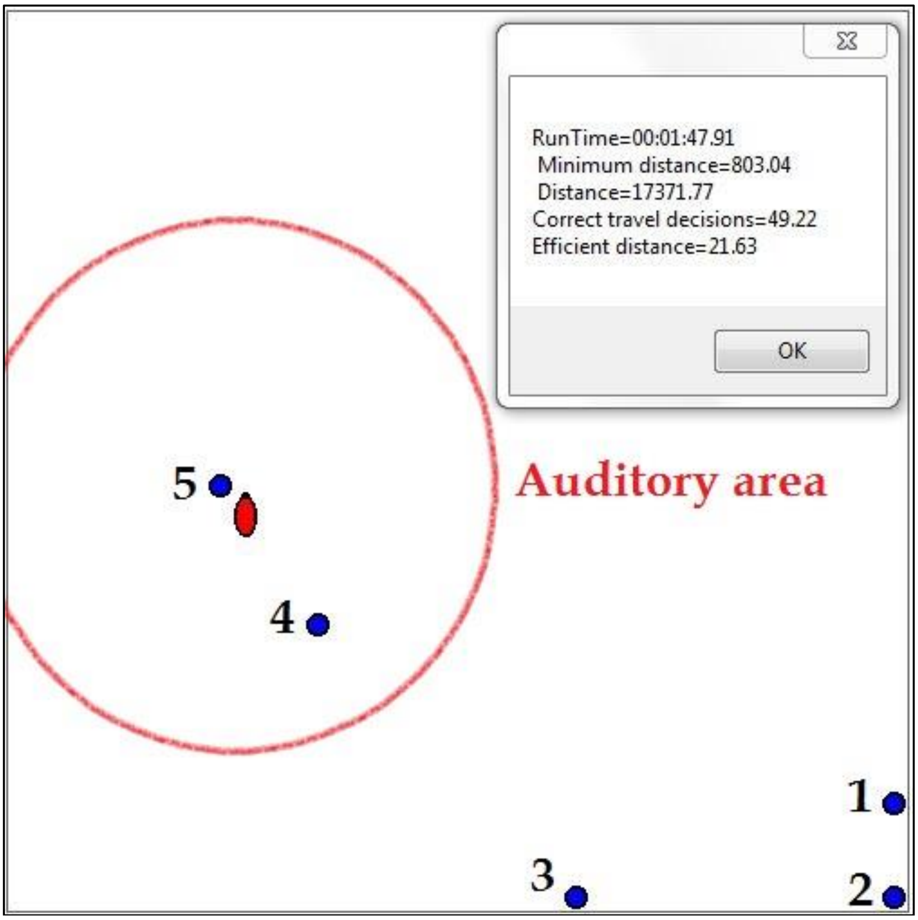

Fig. 6.1. The first audio game

In the second audio game, the 600x600 pixels virtual field has been divided into 5 regions. The upper 3 regions have a size of 200x300 pixels, while the lower 2 regions have a size of 300x300 pixels (Fig. 6.2.). Thus, the upper 3 regions have a smaller search area (with a lower level of difficulty), while the bottom two have a larger search area (with a higher level of difficulty) [59].

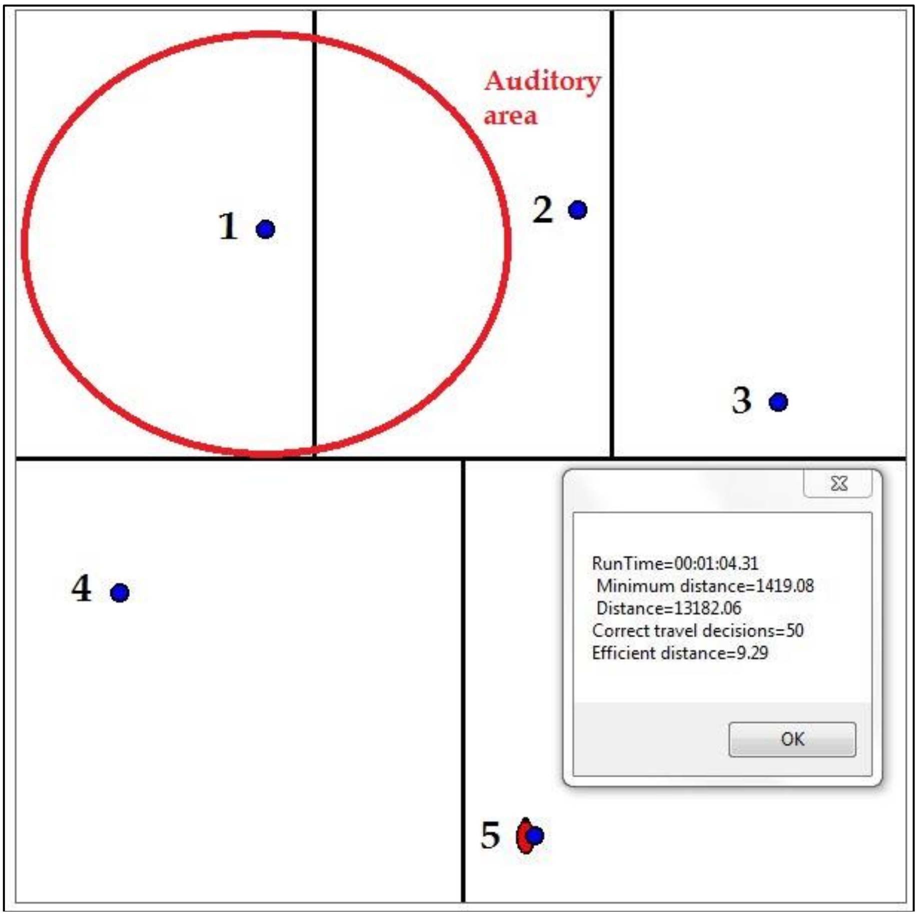

Fig. 6.2. The second audio game

6.1.2. Sound stimuli used in the experiments

The auditory stimuli employed in the games are 3D binaural sounds (delivered via headphones) that have been synthesized using the CSound programming language for sound processing [118] (continuous white noise filtered with the non-individualized HRTFs from the MIT database [117]) and auditory icons that have been used to notify the players in the following situations: when they stop or restart the navigation, discover a target, when they get too close to the borders of the virtual settings or when the game has finished. In both games, the emitting sound sources are positioned at 0 degrees elevation in the vertical plane.

The audio representation is based on the sound intensity encoding of distance between the virtual position of the player and the current location of the sound target (the sound intensity is inversely proportional to the square of distance, so that the perceived amplitude increases continuously as the player gets nearer to the auditory target (detailed in equation 5.2.)) [59].

6.1.3. The experimental procedure

We tested the sound localization ability and spatial representation of 12 sighted subjects who were previously unfamiliar with the perception of 3D binaural sounds. They participated in 2 experiments (the first experiment used the first game described in 6.1.1., while the second experiment employed the audio game with pre-determined regions) that were separated by a time interval of 7 days. Both experiments had 2 test sessions in which the subjects had to complete the games as fast as possible and a 3 minutes training session in which they have been presented different 3D sounds. In order to play the game, the users needed to rely on their auditory localization abilities to create a mental map of the environment and to find the 5 sound-emitting hidden targets. The experiments took place in Ploieşti, Romania, in May-June 2014.

6.1.4. Results

In order to analyze the results of both experiments, we took into account the following parameters:

- **P1: The ratio of the total length of the path travelled by the player to the minimum distance between the 5 targets**. For effective results, this ratio should be as low as possible;
- **P2: The percent of correct travel decisions**, calculated as the ratio of the number of movements towards the target to the total number of movements around it when the sound source was active;
- **P3: The total time (in seconds) needed to complete the game**.

The average results of the experiments are presented in Table 6.1. and Table 6.2.

Table 6.1. The results of the first experiment

Experiment 1			
Parameter	**Session 1**	**Session 2**	**Improvement**
P1	13.06	7.25	44.48%
P2 (%)	57.09	60.28	3.19%
P3 (seconds)	91.75	80.33	12.44%

Table 6.2. The results of the second experiment

Experiment 2			
Parameter	**Session 1**	**Session 2**	**Improvement**
P1	3.33	3.11	6.6%
P2 (%)	61.74	64.78	4.7%
P3 (seconds)	68.75	50.25	26.9%

For the first experiment, the values of parameter P1 significantly improved in the second test session with 44.48% (the results are statistically significant in a Student t-test for dependent means, $t=-2.5$, $p \le 0.05$). However, although there are improvements in the mean values for both parameters - P2 (the mean rate of correct travel decisions increased from 57.09% to 60.28% in the second day of test) and P3 (the mean game completion time decreased from 91.75 seconds to 80.33 seconds), they are not statistically significant at $p \le 0.1$.

Moreover, in the case of the second experiment there is a slight improvement in the mean rate of P1 (6.6%), P2 (4.7%) and a more significant improvement for P3 (26.9%). However, none of these results are statistically significant at $p \le 0.1$.

In what concerns the individual improvements, 66% (respectively, 75% in the second experiment) of the subjects improved the time needed to complete the game from the first to the second session of the experiment, 66% (correspondingly, 58% for the second experiment) took better travelling decisions regarding the position of the target sound sources, while 91% (respectively 66% in the second experiment) of the participants obtained a significant decrease in the value of parameter P1.

In a brief conversation that took place after the second experiment, the subjects concluded that the presented audio games are useful, entertaining and that they can bring great benefits to the enhancement of their auditory and perceptual skills. Also, they considered them to be completely different from the very common video games played before. Our subjects expressed their satisfaction regarding the quality of immersion and high level of interactivity of the proposed games. Also, they would encourage the future development of more advanced versions of these audio games that could be played by both the sighted and the visually impaired people.

6.1.5. Discussion

In both experiments, the participants were able to successfully use 3D auditory cues in order to map the virtual environment and to perform simple exploration tasks. They based their navigation on both the 3D sounds which conveyed spatial information regarding the direction and position of the objects in space, but also on the perception of continuous changes in sound intensity that encoded the distance between the virtual position of the listener and the target sound sources.

The findings of this study demonstrate that the 3D binaural sounds and the inversely proportional encoding of distance can be effectively used for navigating in a virtual auditory environment. Furthermore, continuous modifications in the physical characteristics of the sound can convey relevant information for the listener at the meta-level of perception. In addition to this, even a short period of training (as shown in our experiments) can help the players to obtain better results in the game [59].

6.2. IMPROVING THE AUDIO GAMES PLAYING PERFORMANCE OF THE VISUALLY IMPAIRED PEOPLE THROUGH MULTIMODAL TRAINING

The purpose of this study is to investigate the improvements that occur in the sound localization and navigational audio game playing performances of the visually impaired people as a result of multimodal (auditory and haptic) training. The results of the experiment demonstrate a rapid adaptation of the subjects to the perception of 3D sounds synthesized with non-individualized HRTFs and enhanced sound localization and navigational skills, as a consequence of the training strategy. These improvements are reflected in higher performances in the post-test session of the experiment.

6.2.1. Sound stimuli used in the experiment

The auditory stimuli were continuous 3D binaural sounds synthesized with non-individualized HRTFs taken from the MIT dataset [117], auditory icons (sounds that create an analogy with real-world events and situations) and earcons (abstract, symbolic sounds used to facilitate the players' navigation through the complex content of the game) [92]. All the sounds have been synthesized using the Csound programming language [118].

The location of a hidden auditory target has been sonified using the combination of white and pink noise in varying proportions, as described in 5.2.

The locations of the obstacles have been encoded through an alarm sound that was spatialized using the non-individualized HRTFs from the MIT database, so that the players could identify the direction of both targets and obstacles. The auditory icons used were the alarm sounds aimed at raising the player's awareness in what concerns approaching the obstacles and the sound of a crash (or accident) when he ran into an obstacle. Earcons have been represented by the sound of a click when the player succeeded to identify the position of a target and a bell ringing that announced the end of the game. As the listeners got nearer to a target or an obstacle, the intensity of the sound increased and, on the other hand, as they got farther, the perceived amplitude decreased. The formula that calculates the perceived volume of the target sound source is the following:

$$\text{Gain factor} = \begin{cases} 0, & d > dmax \\ GFMIN + (GFMAX - GFMIN) * (1 - \frac{d}{dmax})^2, & 0 \le d \le dmax \end{cases} \tag{6.1}$$

Where d is the current distance between the position of the listener and the target or obstacle, *dmax* is the maximum recorded distance from the target (for the obstacles, *dmax*=150 pixels), *GFMIN*=0.05 (the minimum gain factor), *GFMAX*=1 (the maximum gain factor) [129].

6.2.2. Applications developed for the experiment

The applications described in this section have been developed in C#, using the Microsoft Visual Studio integrated development environment and employ the sound stimuli presented in subchapter 6.2.1.

THE AUDIO GAME

In the audio game the listeners were required to identify the location of several hidden auditory targets while trying to avoid blocking obstacles (Fig. 6.3.). The game had 10 levels of difficulty designed in the 2D space, with varying numbers of targets and obstacles, as it follows:

- Level 1: 1 target and 1 obstacle
- Level 2: 1 target and 2 obstacles
- Level 3: 2 targets and 2 obstacles
- Level 4: 2 targets and 3 obstacles

- Level 5: 3 targets and 3 obstacles
- Level 6: 3 targets and 4 obstacles
- Level 7: 4 targets and 4 obstacles
- Level 8: 4 targets and 5 obstacles
- Level 9: 5 targets and 5 obstacles
- Level 10: 5 targets and 5 obstacles

At each level, the target sound sources were positioned starting from the bottom border of the playing window to the top of it, so that once the player discovered one target, he was required to look for the next ones up, reducing thus the searching area and further preventing the occurrence of front-back localization misjudgments. Only one sound target was active (could be heard) at a moment of time. When the position of the current target has been identified, it consequently became inactive and the next target (in ascending order, from the bottom side of the playing window to the top of it) became audible. One obstacle became audible when the user got inside a range of 150 pixels radius around it. In the case when the distance between two obstacles was less than 150 pixels and the listener was positioned in their close proximity, he could simultaneously perceive two sound stimuli with different directional cues, corresponding to each of the obstacles. As the subjects were modifying their position in respect with the active target and the surrounding obstacles, they could perceive changes in both the spectral content of the sound, its intensity and its localization. The players were required to navigate freely, using the mouse or touchpad movement as interaction modality. Prior to the experiment, the subjects have been instructed to use the mouse or the touchpad, according to preference or prior experience [129], [130].

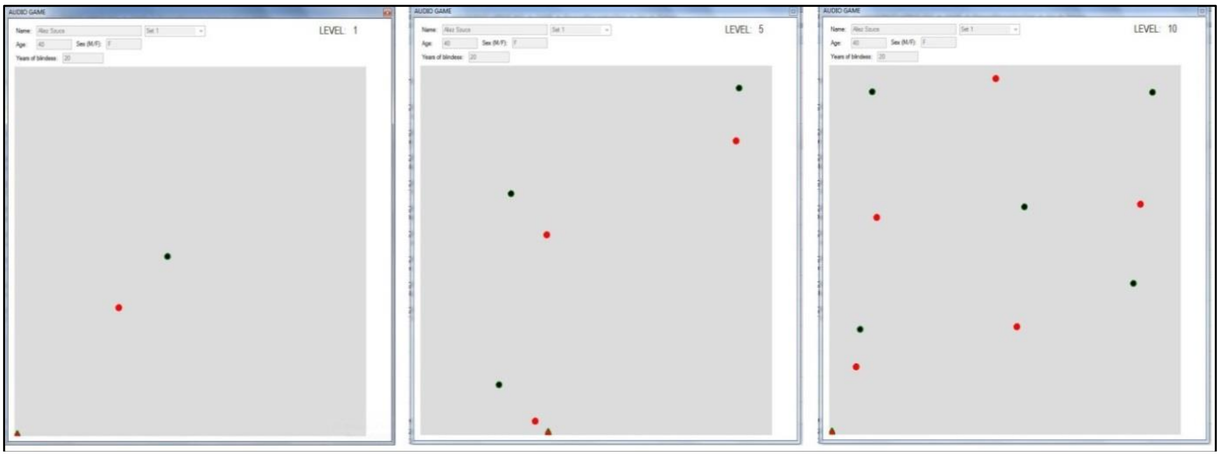

Fig. 6.3. The design of the levels 1, 5 and 10 from set 1 of the audio game. The green circles represent target sounds and the red circles represent obstacles.

THE GAME EDITOR

The experimenter can design the layout of the game (the distribution of auditory target objects and obstacles) in a well-defined and interactive manner. Thus, he is required to introduce the number of the level and to place items (either targets or obstacles) on the canvas (Fig. 6.4.). The positions of the target objects need to be set first (by pressing the “Set Targets” button from the user interface application), in ascending order, from the bottom side of the playing canvas to the top of it. Subsequently, the user is required to press the “Set Obstacles” button in order to start adding obstacles to the current level. In this case, no more than 5 obstacles per level can be introduced. The locations of the target objects or obstacles are set by clicking on the playing window in the positions where the items need to be placed. The game layout components are encoded by green (the target objects), respectively red circles (in the case of the obstacles) [129].

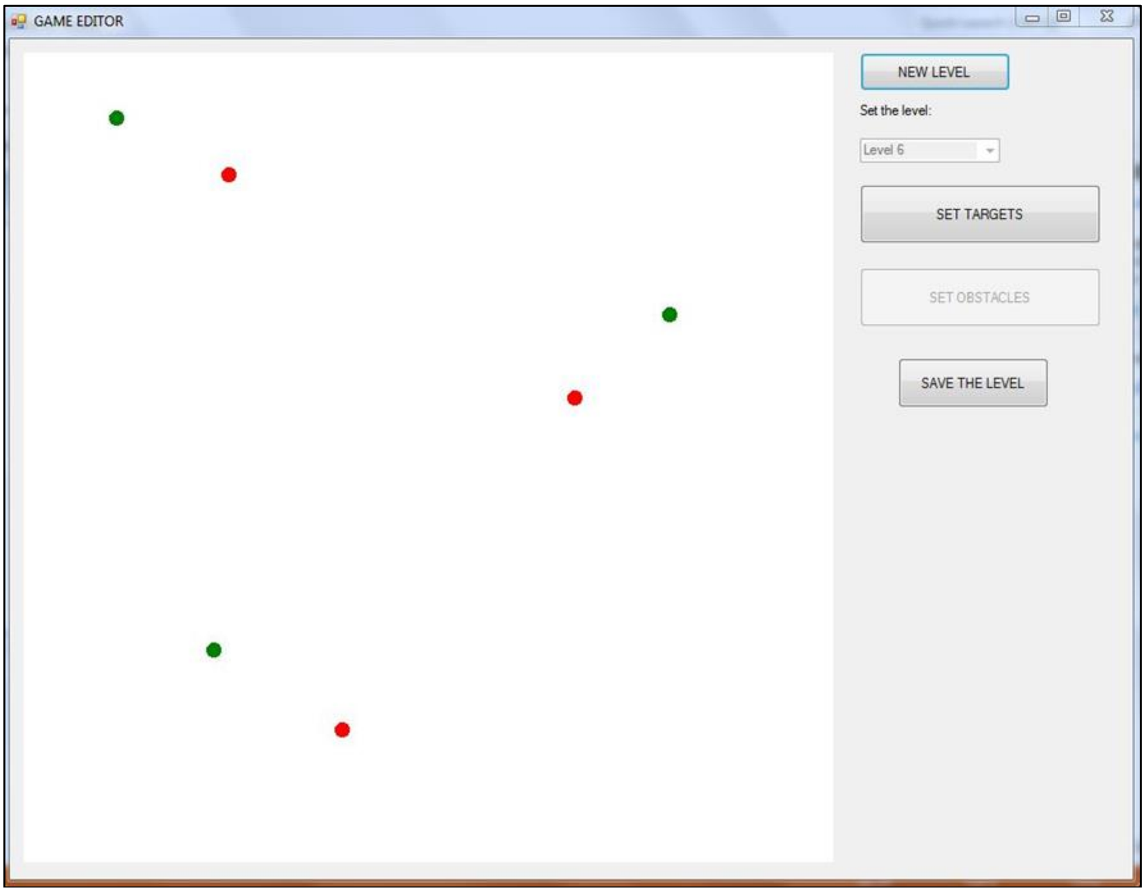

Fig. 6.4. The Game Editor

The data for each level is stored in separate .txt files that comprise on distinct lines the item type ("t" for targets and "o" for obstacles) and the 2D coordinates in the horizontal plane for each of them. For example, the content of the "Level 3.txt" file, where we have 2 target sound sources and 2 obstacles is the following:

Level 3
t 127 481
t 627 36
o 201 559
o 630 125

THE GAME ANALYZER TOOL

The Game Analyzer tool (Fig.6.5.) allows real-time visualization and audio playback of the users' performances for each round (for each target object identification) and level.

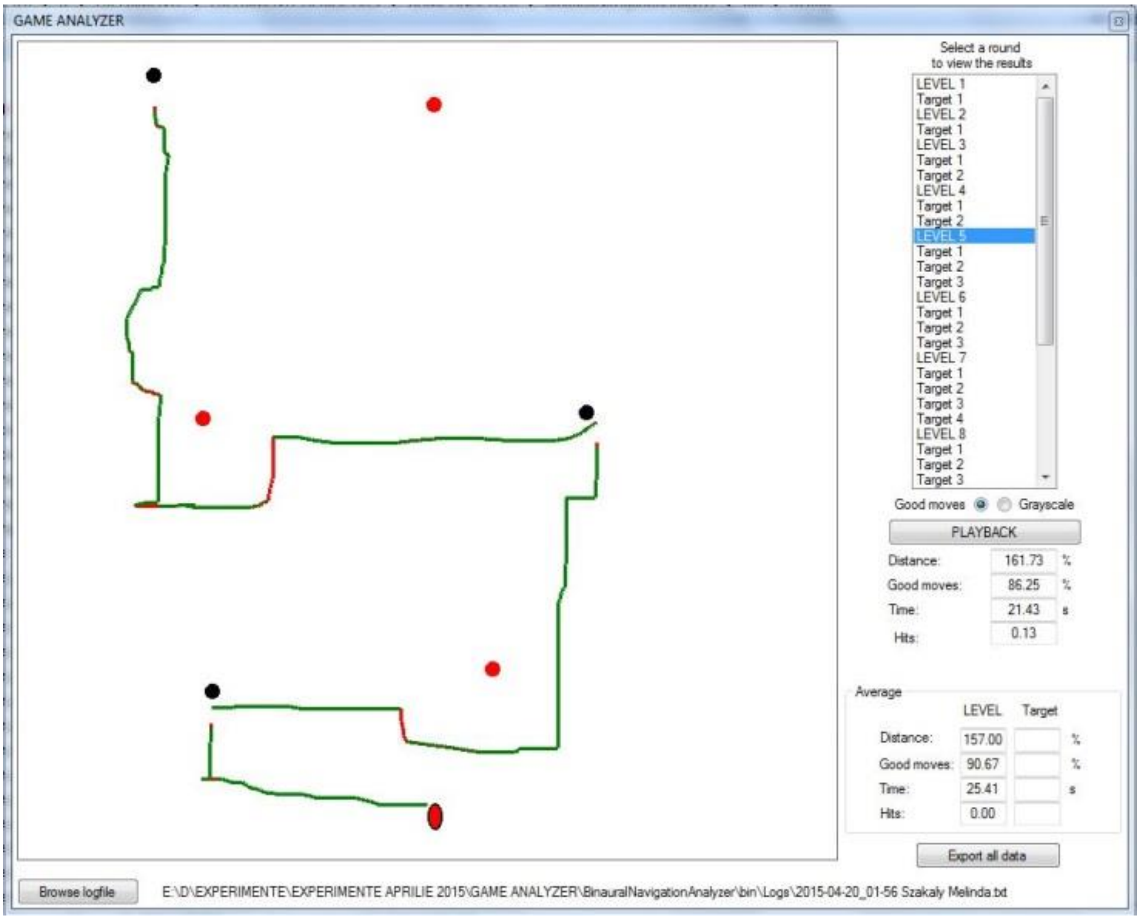

Fig. 6.5. The Binaural Game Analyzer

Fig. 6.5. presents the player's performance for level 5 (which has 3 sound targets and 3 obstacles), using the "Good moves" visualization option.

By pressing the "Browse log file" button, the experimenter can upload the log file that has been created and updated during the gameplay session, storing all the data concerning the user's performance. The experimenter can select a round or a level from the list box control on the right side of the interface. The "Good moves" visualization option displays the player's efficient displacements (mouse movements effectuated towards the sound source) as green colored segments, while the wrong movements that maximize the distance to the sound source are represented by red segments. The "Grayscale" visualization option presents the path travelled by the listener from the starting position until he reaches the target object of each round in progressive grayscale color tones, from light gray to black (Fig. 6.6.). By pressing the "Playback" button, the experimenter can see in real-time the performance of the player and simultaneously hear in the headphones the 3D sounds that the user perceived while navigating from the starting position to the target sound source. Furthermore, the Game Analyzer interface displays the mean values of parameters P1-P4 for each round and level, as well as for all the levels of the game. The "Export all data" button allows the statistical information to be exported to an Excel file, together with the player's personal information, such as name, age, sex and the extent of the visual impairment condition.

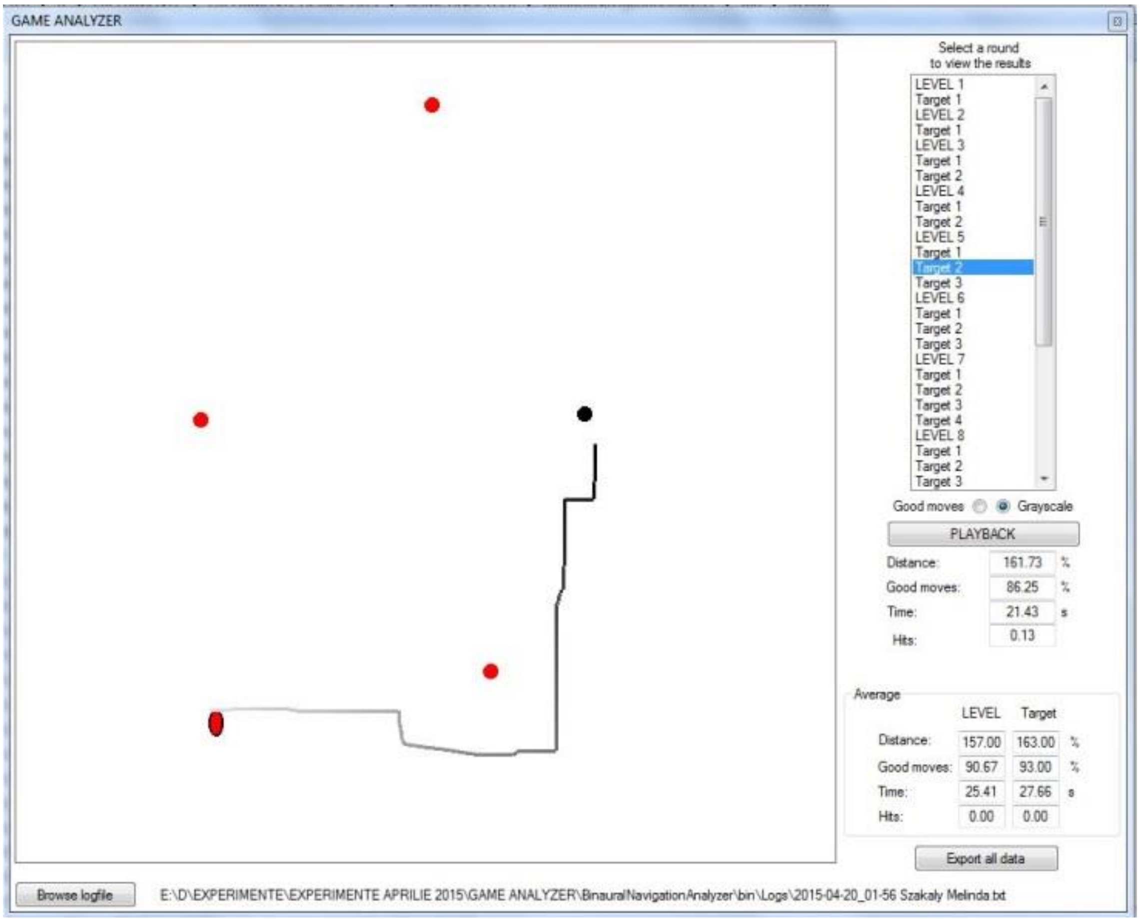

Fig. 6.6. The “Grayscale” visualization option

Fig. 6.6. presents the player’s performance for the second round (when the player discovers the second object of the level) of the 5th level of the game, using the “Grayscale” visualization option which paints the path travelled by the user in progressive grayscale tones.

The purpose of the Game Analyzer application is to evaluate the game playing performances of the users through a complete statistical analysis (the mean values of the four studied parameters, distributed for each round, level and for all the levels of the game), as well as to provide a deeper understanding of the issues the subjects are confronted with during gameplay – the perception of 3D sounds, sound localization accuracy, front-back confusions, auditory-based spatial navigation and interaction modality [129], [130].

6.2.3. The sound localization experiment

This study comprises the assessment of the sound localization skills of 10 visually impaired subjects who were required to play the audio game with hierarchical levels of difficulty described in 6.2.2. In the game, the players had to identify the location of several hidden auditory targets while trying to avoid the obstacles standing in their way. The experiment was composed of a pre-test session (in which the subjects were asked to play the game twice, for two different sets of levels), a training session (aimed at helping the visually impaired individuals to adapt to the perception of 3D sounds through multimodal interaction – both auditory and haptic, using the haptic belt presented in 5.1.2., with the only difference that in this case there have been used 24 vibration motors, placed at 15 degrees difference around the subject’s head) and a post-test session (similar in structure and difficulty to the pre-test), in which the degree of sound localization improvement following training has been evaluated. The experiment took place in April 2015 in Györ, Hungary [129].

TARGET GROUP

10 visually impaired individuals (5 women and 5 men, living in Györ, Hungary, aged 27-63, mean age = 43, with a percent of residual vision ranging between 0% and 15%. Two of the subjects were congenitally blind, one was congenitally visually impaired - with a percent of vision of 10%, while the others were late-onset visually impaired) took part in the experiment. Previous to the start of the tests, the subjects were informed about the purpose of the experiment and gave their full written consent to participate in it.

THE PRE-TEST AND POST-TEST SESSIONS OF THE EXPERIMENT

During the pre-test session, the subjects were required to play the game twice, with different sets of levels (set 1 and set 2) (Fig. 6.7.). Before the pre-test, the subjects were presented the purpose of the game and the main auditory cues used in the sonification strategy. Moreover, in order to allow them to get familiarized with the aim of the game and with the perception of 3D sounds, the subjects were allowed to practice playing the game as long as they considered it was necessary prior to the start of the tests. Usually, most of the subjects played the game once (10 levels), for an average time of 10 minutes, until they got accustomed to the sonification approach and the aim of the game.

The studied parameters were:

- P1: The ratio of the distance travelled by the player (from the starting position until he discovers the location of the current target) to the minimum possible distance (the Euclidean distance between the starting point and the position of the current target). For the first target of any level, the starting position was the center of the bottom border of the playing window. For the other targets, the starting position was the location of the previously identified target;
- P2: The percent of correct travel decisions, defined as movements effectuated towards the sound source (minimizing the distance between the user’s virtual location and the position of the target);
- P3: The mean level completion time (in seconds);
- P4: The mean number of obstacle hits.

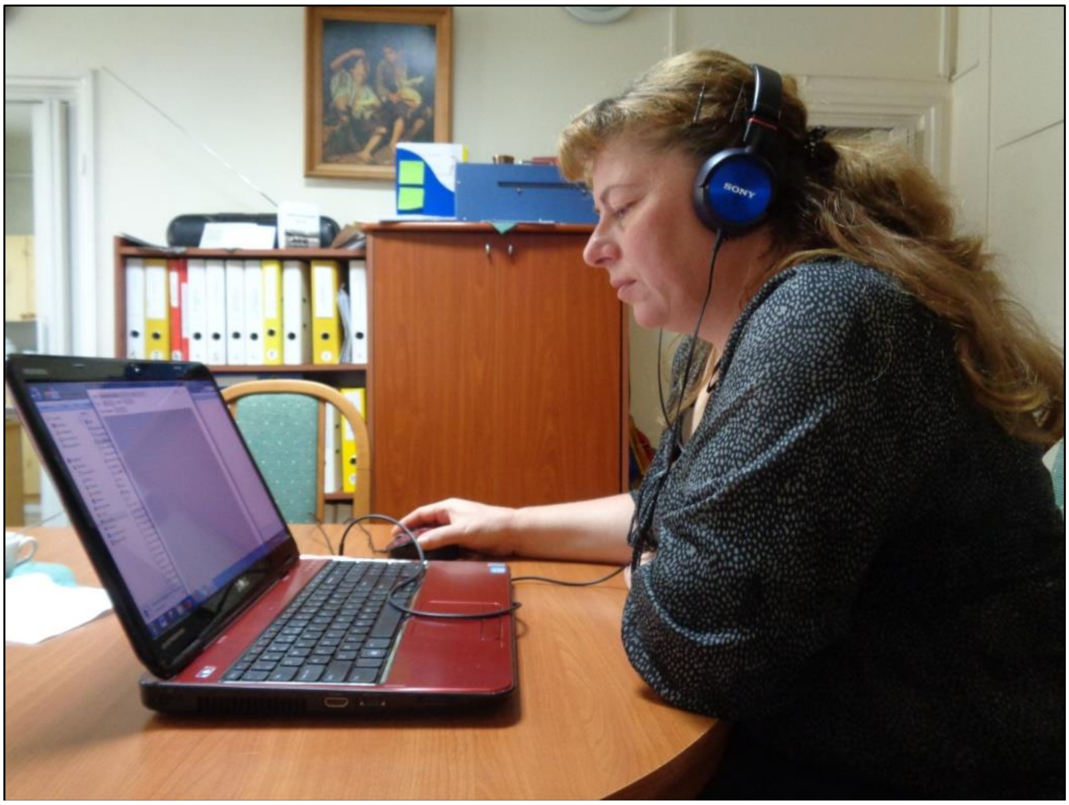

Fig. 6.7. Visually impaired subject during gameplay

The post-test session, that took place one day after training, has been carried out in exactly the same conditions as the pre-test session, using the audio-based game. The goal of the post-test session was to assess the level of sound localization and navigational skills improvement achieved after training [129].

THE TRAINING SESSION OF THE EXPERIMENT

During the training session, the visually impaired subjects were offered haptic feedback in what concerns the directionality of the sound source. Thus, they were required to wear stereophonic headphones and the haptic belt described in 5.1.2. - containing 24 vibration motors (placed at 15 degrees difference around the head) - that transmitted vibrations corresponding to the direction of the sound source in space.

The training session took place in two consecutive days. In each day, the subjects have undergone three training blocks. Each training block had a duration of 3 minutes and required the subjects to listen to a series of 24 sounds (emulated in clockwise order from 0 to 345 degrees) and then to randomly generated auditory stimuli with a duration of 4 seconds. The sounds that have been used were the combination of white and pink noise in varying proportions - the same encoding used for the game described in the previous section. Each sound stimulus perceived through the headphones has been accompanied by a train of 4 vibrations on the haptic headband (one vibration per second) corresponding to the direction of the sound in space. The purpose of the training session was to help the visually impaired subjects to get used to the perception of 3D sounds and to create an effective crossmodal association (haptic and auditory) that would enable them to easily identify the direction of the 3D sounds. The high resolution of the haptic belt (where the vibration motors have been placed at 15 degrees difference all around the subject's head) allowed for more accurate training. (Fig. 6.8.) [129].

Fig. 6.8. Visually impaired subject during the training session

6.2.4. Results

RESULTS OF THE PRE-TEST AND POST-TEST SESSIONS

Table 6.3. briefly presents a statistical overview of the results obtained by the subjects (mean, SD – standard deviation, minimum and maximum value) in both the pre-test and post-test sessions of the experiment.

Table 6.3. A brief statistics of the results

	Pre-test session				Post-test session			
	P1 (distance ratio)	**P2 (% correct moves)**	**P3 (seconds)**	**P4 (obstacle hits)**	**P1 (distance ratio)**	**P2 (% correct moves)**	**P3 (seconds)**	**P4 (obstacle hits)**
Mean	4.2	70.6	41.6	0.4	3.0	77.4	27.7	0.2
SD	1.8	6.6	10.4	0.3	1.5	8.1	17.5	0.2
Min	1.8	58.5	23.5	0.1	1.5	66.2	10.9	0.0
Max	8.6	83.1	61.9	1.3	6.4	90.0	60.7	0.5

The results show that the subjects recorded substantial improvements in the post-test session of the experiment for all the four studied parameters. There is a significant improvement for parameter P2 (the results are statistically significant in an ANOVA test at $p \leq 0.1$ and in a Student t-test for dependent means where $t=1.51$, $p \leq 0.1$), P3 (in an ANOVA test at $p \leq 0.05$ and in a Student t-test where $t=-3.42$ at $p \leq 0.05$) and P4 (in an ANOVA test at p≤0.1). The mean rate of P1 decreased with 27% in the post-test session of the experiment, while the percent of correct travel decisions towards the target source increased with approximately 7% (from 70.6% to 77.4%). Moreover, the average time to complete a level (parameter P3) reduced with 33% (from 41.6 to 27.7 seconds) and the mean number of obstacle hits per level decreased with 50% (from 0.4 to 0.2). In addition, more than 80% of the subjects succeeded to enhance their sound localization and navigational performance in the post-test session of the experiment (Table 6.4.). It should be emphasized that all the participants recorded a higher rate of correct travel decisions towards the target sound source, demonstrating that the training procedure helped them to recalibrate their spatial hearing and improved the perception of 3D binaural sounds synthesized using non-individualized HRTFs.

Table 6.4. Improvements recorded in the post-test session

	P1 (distance ratio)	**P2 (% correct moves)**	**P3 (seconds)**	**P4 (obstacle hits)**
Average improvement in the post-test session	27%	6.81%	33.3%	50.3%
Percentage of subjects with improvements in the post-test session	90%	100%	90%	80%

Table 6.5. presents the mean results for all the four parameters, for both sets of levels, in the pre-test and post-test sessions of the experiment. For the first set of game levels, there have been recorded significant improvements for parameters P1 (the differences are statistically significant in an ANOVA test at p≤0.1 and in a Student t-test, t=-3.46, p≤0.05) (Figure 6.9.), P2 (in an ANOVA test at p≤0.1 and in a Student t-test for dependent means where t=3.24, p≤0.05) (Figure 6.10.), P3 (the results are statistically significant in an ANOVA test at p≤0.1 and in a Student t-test where t=-3.55 at p≤0.05) (Figure 6.11.) and P4 (in an ANOVA test at p≤0.1 and in a Student t-test where t=-2.01 at p≤0.1) (Figure 6.12.). The mean rate of P1 decreased with 30.9% (from 4.3 to 2.9), the percent of correct travel decisions increased with 7.2% (from 68.6% to 75.8%) and the mean completion time per level reduced with 34.9% (from 48.1 seconds to 31.3 seconds). Similarly, the mean number of hits per level (parameter P4) decreased with 49.2% (from 0.35 to 0.18 in the post-test session).

In what concerns the results obtained for the second set of levels, the results obtained in both the pre-test and in the post-test sessions are higher than those recorded for the first set. Thus, the mean value of parameter P1 decreased with 14% (from 3.1 to 2.7 – although the results are not statistically significant), the mean rate of correct travel decisions increased with 7.7% (from 74.3% to 82.1% - the differences between the performance in the pre-test and post-test sessions are statistically significant in an ANOVA test at p≤0.05 and in a Student t-test where t=3.4 at p≤0.05), the level completion time reduced with 20.4% (from 32.5 seconds to 25.9 seconds in the post-test session – the differences are not statistically significant), while the mean rate of obstacle hits per level decreased with 14.2% (from 0.30 to 0.26 – the results are not statistically significant) (Table 6.6.).

Table 6.5. Results for the first and the second set of levels

	Pre-test session				Post-test session			
	P1	**P2 (%)**	**P3 (seconds)**	**P4**	**P1**	**P2 (%)**	**P3 (seconds)**	**P4**
Set 1	4.3	68.6	48.1	0.35	2.9	75.8	31.3	0.18
Set 2	3.1	74.3	32.5	0.30	2.7	82.1	25.9	0.26

Table 6.6. Improvements in the post-test session of the experiment for both sets of levels

Percentage of improvement in the post-test session	**P1**	**P2**	**P3**	**P4**
Set 1	30.9%	7.2%	34.9%	49.2%
Set 2	14%	7.7%	20.4%	14.2%

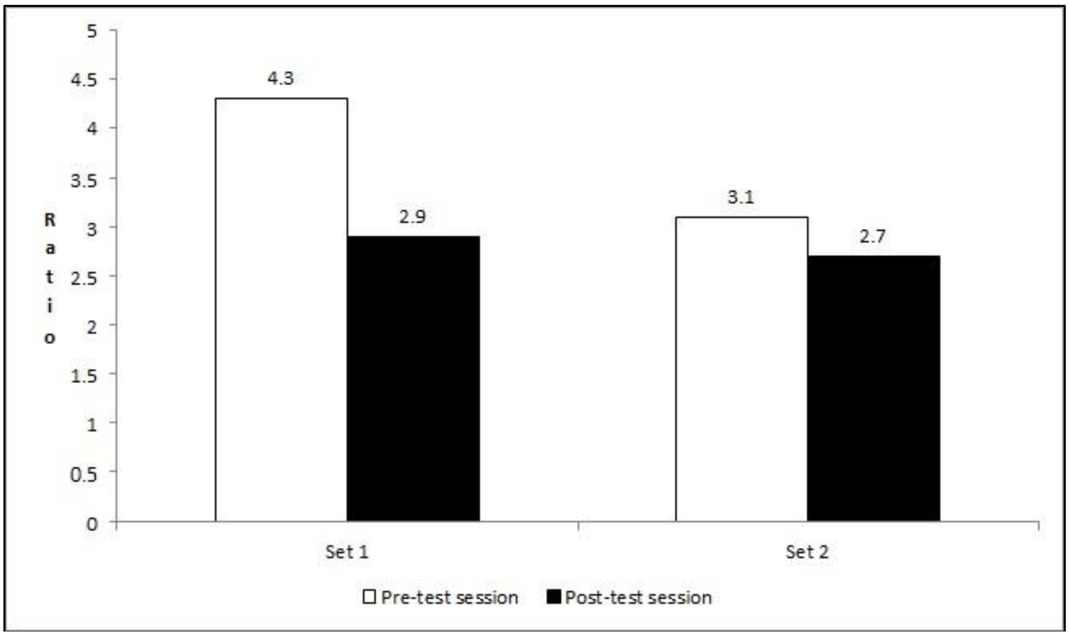

Fig. 6.9. Evolution of parameter P1 in the pre-test and post-test sessions of the experiment, for both sets of levels

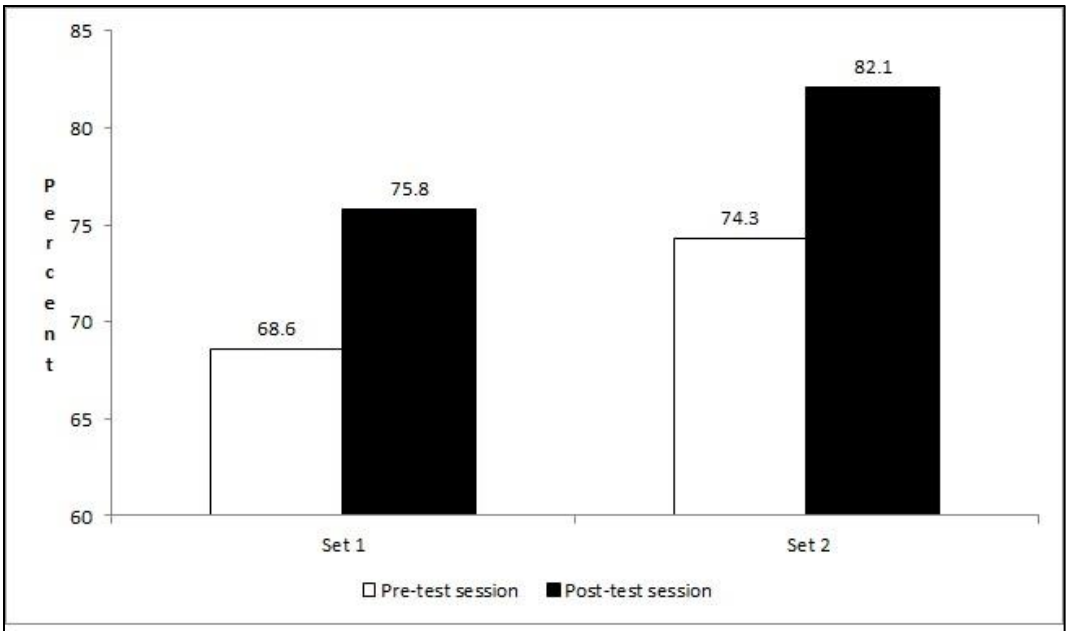

Fig. 6.10. Evolution of parameter P2 in the pre-test and post-test sessions of the experiment, for both sets of levels

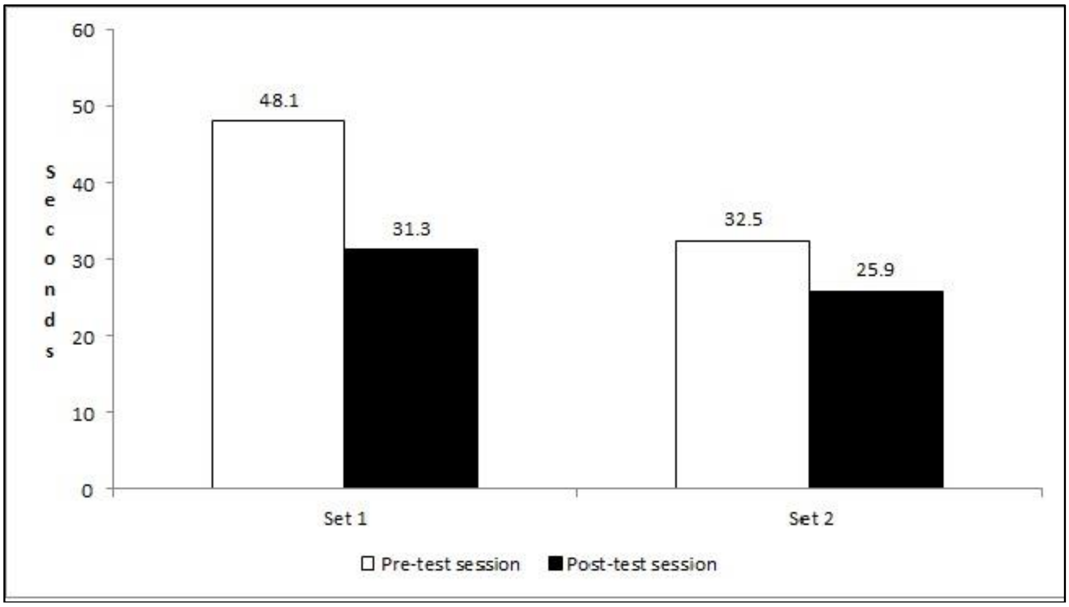

Fig. 6.11. Evolution of parameter P3 in the pre-test and post-test sessions of the experiment, for both sets of levels

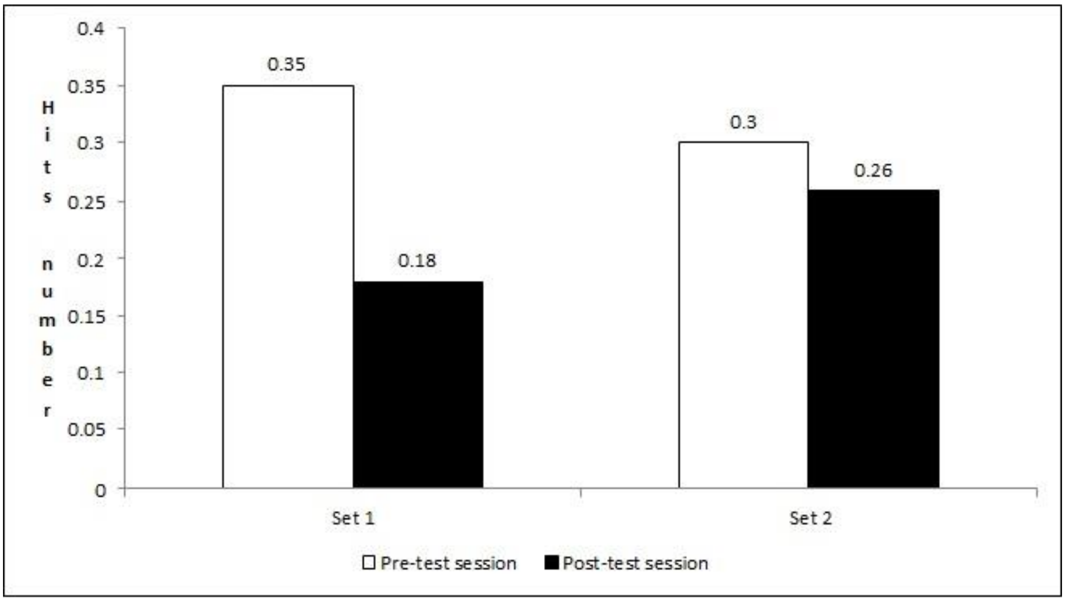

Fig. 6.12. Evolution of parameter P4 in the pre-test and post-test sessions of the experiment, for both sets of levels

7 of our visually impaired subjects have participated in the sound localization experiment presented in chapter 5.

The results presented in Table 6.7. show that the mean rate of parameters P1 and P2 is higher in the current experiment than in the previous one, demonstrating the persistence of the spatial auditory adaptation of the human hearing system and its continuous improvement in time.

Table 6.7. Comparative analysis between the results of the post-test session of the previous experiment and the results of the current one

SESSION	P1	P2 (%)
Post-test session of the previous experiment	6.6	66.1
Pre-test session of the current experiment	3.8	71.7
Post-test session of the current experiment	2.84	80

The mean rate of parameter P1 improved with 43.1% ($t=1.68$, $p \le 0.1$) and that of parameter P2 with 5.64% (the results are statistically in an ANOVA test at $p \le 0.1$ and in a Student t-test where $t=2.08$ at $p \le 0.1$) between the post-test session of the previous experiment and the pre-test session of the current one (Figure 6.13., Figure 6.14.). Also, 85% (6 out of the 7 subjects) recorded more accurate sound localization abilities in the pre-test session of the current experiment that in the post-test phase of the previous one, fact that can be explained by their familiarity with the tasks and with the sound stimuli - the perception of directionality conveyed by the 3D sounds, the spectral characteristics of the white and pink noise and the continuous change in sound intensity that is inversely proportional to the distance between the listener and the sound source.

The final results demonstrate significant improvements in the post-test session of the present experiment, surpassing those obtained in the previous experiment with 57.4% (the results are statistically significant in an ANOVA test at $p \le 0.1$) for parameter P1 (from 6.68 to

2.84) and with 13.8% (the differences are statistically significant in an ANOVA test at $p \leq 0.05$ and in a Student t-test for dependent means where $t=3.5$ at $p \leq 0.5$) for P2 (from 66.15% to 80%).

Moreover, the players who participated in the previous experiment were more proficient in accomplishing the navigational and sound localization tasks required in the current game than their inexperienced counterparts. At the same time, the experienced users recorded a higher improvement in the post-test session of the current experiment (compared to the pre-test session), increasing for their percentage of correct travel decisions with 8.3% (from 71.7% to 80% - the results are statistically significant in an ANOVA test at $p \leq 0.05$ and in a Student t-test where $t=3.66$ at $p \leq 0.05$) [129].

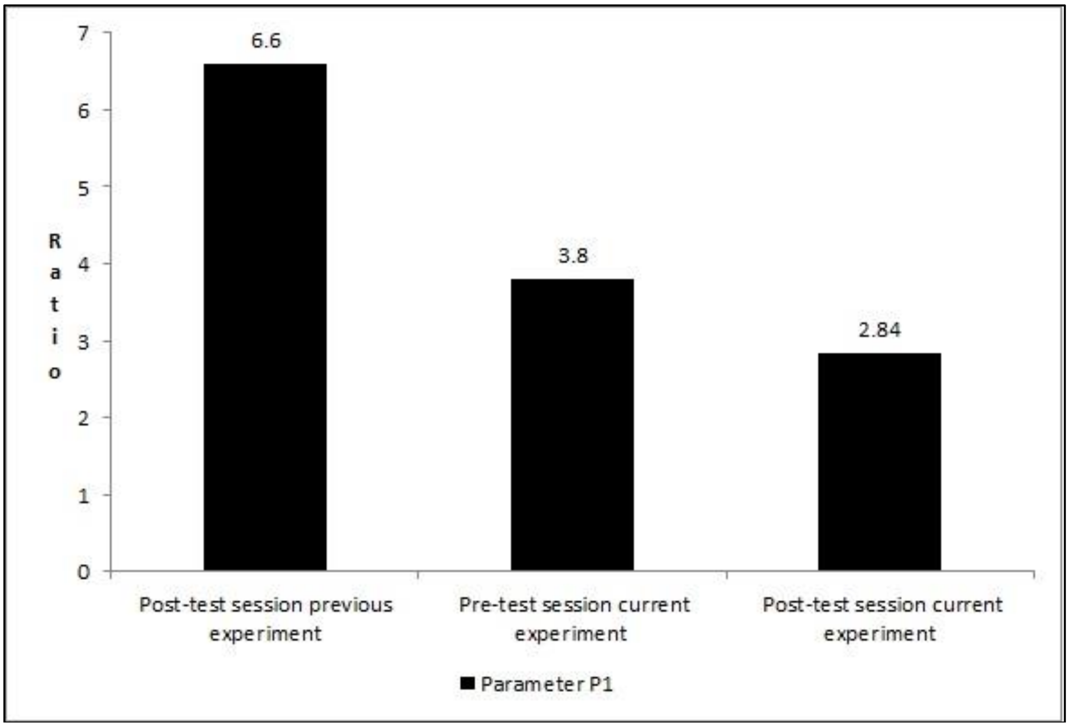

Fig. 6.13. Evolution of parameter P1, compared to the post-test session of the previous experiment

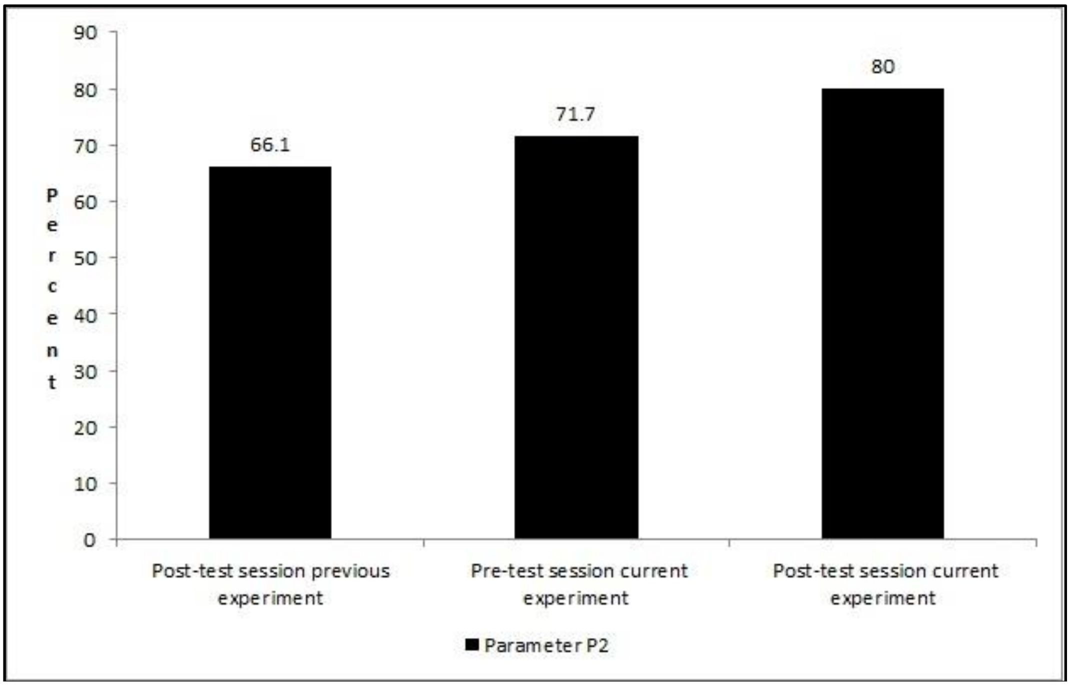

Fig. 6.14. Evolution of parameter P2, compared to the post-test session of the previous experiment

By placing the vibration motors at 15 degrees difference on the haptic belt, the improvements achieved by the visually impaired subjects in the post-test session of the experiment presented in this chapter are higher than those obtained in the post-test phase of the experiment described in chapter 5, for parameters P1 (3, respectively 6.8) and P2 (77.4%, respectively 66.1%), but not for P3 (27.7 seconds, compared to 22 seconds).

RESULTS OF THE USABILITY STUDY

After the post-test session, the subjects were asked to complete a usability questionnaire (10 questions on the 5-point Likert scale), in order to express their opinions about the experiment:

1. How often do you use the headphones for the computer or for the mobile phone?
a. Very often b. Often c. Neither often, nor rarely d. Rarely e. Very rarely

2. In the audio game, I could detect the location of the target sounds:
a. Very easily b. Easily c. Neither easily, nor with difficulty d. With difficulty e. With very much difficulty

3. In the audio game, I could detect the location of the target sounds situated in the front and in the back:
a. Very easily b. Easily c. Neither easily, nor with difficulty d. With difficulty e. With very much difficulty

4. In the audio game, I could detect the location of the obstacle sounds:
a. Very easily b. Easily c. Neither easily, nor with difficulty d. With difficulty e. With very much difficulty

5. I consider that playing the audio game helped me to improve my localization skills:

a. Very much b. Much c. Neither much, nor little d. Little e. Very little

6. I consider that the audio game is motivating and interesting:
a. Very much b. Much c. Neither much, nor little d. Little e. Very little

7. I consider that this audio game would be effective for training the visually impaired people to learn how to use an assistive device :
a. Very much b. Much c. Neither much, nor little d. Little e. Very little

8. The game is intuitive and I have learned to use it quite quickly:
a. Very much b. Much c. Neither much, nor little d. Little e. Very little

9. I would like to play the audio game for training my sound localization abilities in the future:
a. Definitely yes b. Yes c. I don't know d. No e. Definitely no

10. I consider that the haptic training (using vibrations) helped me to improve my sound localization skills by providing auditory and haptic feedback:
a. Very much b. Much c. Neither much, nor little d. Little e. Very little

70% of the subjects considered that they could detect the location of the target sounds very easily or easily and 80% stated that they did not encounter difficulties in identifying the location of the sound sources situated in the front or in the rear. Moreover, 90% considered that they could detect the location of the obstacles very easily or easily. At the same time, 90% of the visually impaired participants appreciated that playing the audio game enhanced their sound localization abilities, that the audio game was motivating and interesting and that it would be an effective training strategy towards learning how to use an assistive device. Furthermore, 70% of the subjects considered that the game was intuitive and easy or very easy to be learnt and used. All the participants unanimously admitted that they would like to play the audio game in the future for training their sound localization abilities. The last question was related to the efficiency of using the haptic belt as a training method. 80% of the subjects appreciated that the multimodal training strategy (based on both auditory and haptic feedback) helped them to improve their sound localization abilities and navigational skills (Fig. 6.15.).

In a short interview that took place after the experiment, the subjects made suggestions and commentaries about the proposed game. They appreciated the application and considered it to be a useful tool for both training and entertainment. At the same time, they suggested that it would be a good idea to develop an official version of the game that could be downloaded or played on the Internet [129].

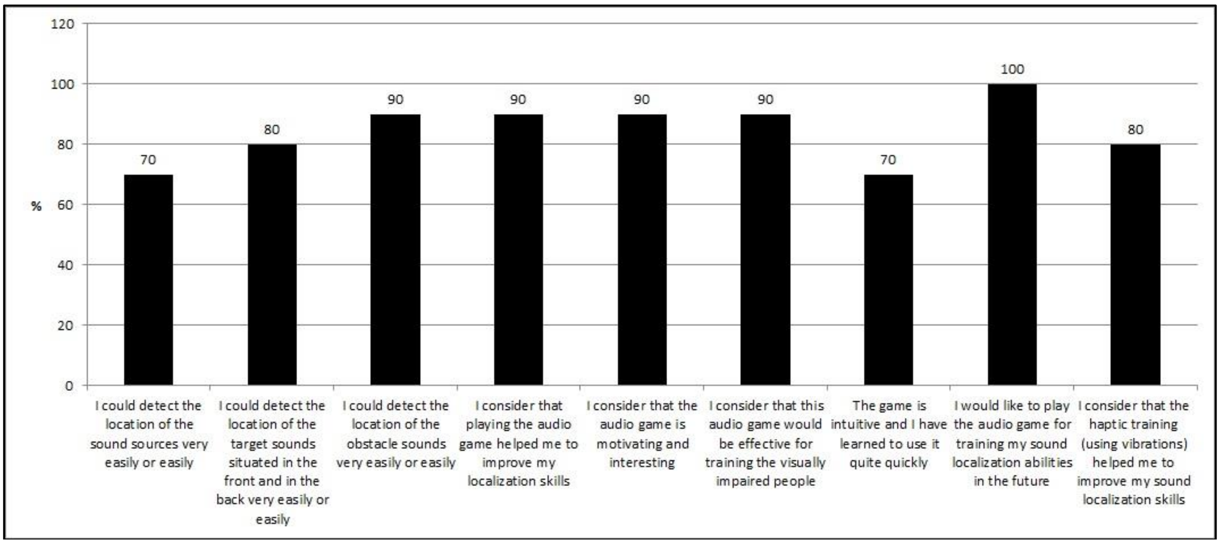

Fig. 6.15. Results of the usability study for the audio game experiments

6.2.5. Discussion

The results of our experiment showed that the visually impaired people succeeded to achieve a rapid improvement of their sound localization abilities and auditory-based navigational skills. These improvements are the result of the multimodal (auditory and haptic) training strategy which demonstrated that learning plays an important role in recalibrating the spatial auditory resolution of the visually impaired individuals. The training procedure helped the subjects to adapt to stimulus conditions that present a mismatch between the spectral cues and the sound direction, such as the use of 3D binaural sounds synthesized with non-individualized HRTFs in virtual auditory environments. Moreover, the perceptual training enabled the subjects to map the virtual settings and to perform simple navigational tasks (target localization, obstacle avoidance and spatial orientation).

The subjects relied their game playing strategy on the perception of directional binaural sounds which gave clear clues about the location of both targets and obstacles in space, but also on the perception of continuous changes in sound intensity. The improvements achieved in the post-test session of the experiment demonstrate that the visually impaired participants have been able to transform the auditory cues into a solid mental representation of the environment. Furthermore, due to the rapid recalibration to the perception of 3D sounds, we consider that other cognitive abilities, such as attention and concentration have been employed. Attention and focus were highly required for associating the auditory and haptic cues that the subjects received during the training session.

The improvement in sound localization accuracy is explained by the perceptual training method that used broadband noises containing more spectral cues for the learning/retrieving process. Moreover, even if these stimuli are not natural, they are effective for training due to their enhanced externalization features. Another argument that supports the efficiency of the training session is that the subjects have been trained only for 24 virtual sound source positions, whereas error reductions and improved spatial perception have been recorded for many other stimuli directions, including the untrained positions. This fact demonstrates that indeed there occurred a spatial recalibration to altered head-related cues and not only a conscious learning process of the spectro-temporal features of the sound. In addition to this, one of the most remarkable results reported in our study are the long-lasting effects of the training sessions performed in the previous experiment. However, the results of the pre-test session of the current experiment are more accurate than those recorded in the post-test phase of the previous one, demonstrating that the spatial auditory remapping is a continuous process and that a new, solid and persistent head model is developed for localizing altered sound cues, in accordance to Hofman's theory [35], [56].

To sum up, the ratio of the total distance travelled by the listeners to the minimum possible distance decreased by 27%, the rate of correct travel decisions towards identifying the location of the target sound sources improved with approximately 7%, the mean level completion time reduced with 33%, while the number of obstacle hits per level decreased by a half in the post-test session of the experiment.

As the results obtained in both the pre-test and post-test sessions of the current experiment are higher than those achieved in the previous one, we demonstrated that the human auditory system is able to continuously improve its sound localization abilities and that experience-driven learning plays a fundamental role in enhancing the navigational and spatial cognitive skills of the visually impaired individuals [129].

6.2.6. Conclusions

The experimental results of this research demonstrated that the visually impaired people are able to perform route-navigational tasks (such as searching for auditory targets or avoiding obstacles) in virtual reality environments using 3D binaural sounds as the only means for orientation. Moreover, the brief multimodal (auditory and haptic) training session helped the subjects to adapt to altered hearing conditions (such as the use of 3D sounds filtered with non-individualized HRTFs) by creating a strong association between the auditory stimuli perceived in the headphones and the vibrations corresponding to the direction of the sound source on the haptic belt. The subjects enhanced their spatial auditory resolution, navigational skills, orientation and mobility and decision making abilities. Moreover, they succeeded to build a solid spatial representation of the virtual environment that conducted to significant improvements in the game performance.

Thus, we conclude that the currently presented game is a useful rehabilitation tool for improving the navigational abilities of the visually impaired people in a safe and reliable environment (such as the virtual one), before using a sensory substitution device that would provide a complete representation of the environment (as the one we intend to develop [131]) under real-world conditions. Furthermore, as the audio games are not restricted to the blind community only, we consider that this game can be played by the sighted people who want to try an alternative to the video games, entertain and train their sound localization skills at the same time [129], [130].

7. CONCLUSIONS AND FUTURE WORK

The experiments presented in this thesis have been conducted in Ploieşti, Romania (with sighted subjects) and in Györ, Hungary (with visually impaired participants).

7.1. THE ORIGINAL CONTRIBUTIONS OF THE THESIS

1. Contributions to the study of 3D binaural sound localization

These contributions reside in the development of an experimental strategy (consisting in a pre-test, a multimodal perceptual feedback based training and a post-test session) that uses two different types of stimuli: narrowband (the "ding" sound) and broadband, more complex auditory cues, such as the combination of white and pink noises in varying proportions according to the direction of the sound source in space.

Various experiments demonstrated that the multimodal perceptual feedback training is more efficient than the unimodal feedback. For instance, in Strelnikov's experiment [136], the group of subjects who received auditory-only feedback obtained the lowest rate of improvement, much lower than the visual-auditory trained group. Using visual-auditory feedback, Zahorik [55] recorded a significant decrease in the front-back confusion rate in the post-test session of the experiment, from 38% to 23%. Important contributions to the study of multimodal perceptual feedback based training (auditory, visual and proprioceptive) belong to Majdak et al [60], Parseihian & Katz [42], Honda et al [39] and Blum [38].

The purpose of this research was to investigate the degree of training-based adaptation, sound localization improvement and front-back confusion reduction for these two types of sounds that present a great potential for being employed in virtual auditory displays, audio games, assistive devices or audio-based systems. For instance, the narrowband sounds can be used as 3D auditory icons (to provide directional information in audio games), auditory emoticons or musicons, to symbolize events, conceptualize data and to increase the level of immersion in virtual auditory environments. On the other hand, the broadband sounds can be integrated into the sonification technique employed for encoding directional information in sensory substitution devices.

The results of this study have been published in the Proceedings of the 18th International Conference on System Theory, Control and Computing, 2014 [59], Proceedings of the International Conferences on Auditory Display, Graz, Austria, 2015 [107], [108], Proceedings of the Interfaces and Human Computer Interaction Conference, Gran Canaria, Spain, 2015 [111], Archives of Acoustics journal (2 papers, currently under review [109], [110]) and in the Journal of Visual Impairment & Blindness [129] (currently under review).

2. The development of the software applications used for the 3D binaural sound localization experiments

Another contribution is represented by the development of the Binaural Navigation Test application (described previously in 5.3.1.) and the applications used during the training sessions of the experiments presented in 5.3.2. and 5.3.3. (The Visual-Auditory Perceptual Training Application and The Haptic-Auditory Feedback Training Application). Furthermore, for assessing the results of the pre-test and post-test sessions, we have developed the Binaural Navigation Analyzer tool, described in 5.4.

3. The multimodal perceptual feedback based training experiments

These experiments have been conducted in Ploieşti, Romania (with 9 sighted subjects) and in Györ, Hungary (with 9 visually impaired subjects). The applications used in these experiments were the Binaural Navigation tool (in the pre-test and post-test sessions of the experiment), the Visual-Auditory Perceptual Training Application (used in the training session of the experiment with sighted subjects), the Haptic-Auditory Feedback Training Application (used in the training session of the experiment with visually impaired subjects) and the Binaural Navigation Analyzer tool used for evaluating the subjects' sound localization performances in the pre-test and post-test sessions of the experiment.

These experiments demonstrated that the human auditory system is able to quickly adapt to altered hearing conditions, such as listening to 3D binaural sounds filtered with non-individualized HRTFs. The adaptation process was the result of a crossmodal perceptual association between the hearing and the visual (in what concerns the sighted subjects) or tactile sense (being the case of the visually impaired participants). Both the sighted and the visually impaired listeners succeeded to improve their sound localization performance by reducing the reversal and precision error rates for both types of stimuli (with significant better results for the rounds that employed the synthesis of white and pink noise in varying proportions, according to the direction of the sound source in space). Although the sighted subjects outperformed their visually impaired counterparts in most of the required tasks, the level of improvement of the visually impaired participants in the second day of training, respectively in the post-test session of the experiment was generally higher than that of the sighted individuals. Also, the percent of subjects who recorded a significant improvement as a result of training is higher in the visually impaired group.

The results of this study have been published in the Proceedings of the 18^{th} International Conference on System Theory, Control and Computing, 2014 [59], Proceedings of the International Conferences on Auditory Display, Graz, Austria, 2015 [107], [108], Proceedings of the Interfaces and Human Computer Interaction Conference, Gran Canaria, Spain, 2015 [111], Archives of Acoustics journal (2 papers, currently under review [109], [110]).

4. **The multimodal (auditory and haptic) training approach** that uses the haptic belt (presented in 5.1.2.) to convey perceptual information concerning the direction of the 3D sound source in space. This training strategy enabled the visually impaired subjects to improve their sound localization performance (they recorded lower angular precision errors and front-back confusion rates after the training session), navigational skills and directional decision-making abilities.

It has been observed that training plays a fundamental role for auditory adaptation to altered hearing conditions and that a rapid adaptation of the auditory system to non-individualized HRTFs is possible through a spatial map recalibration with multimodal sensory associations. The proposed method will be used to train the spatial auditory resolution of the users of an assistive device that aims to provide a rich representation of the environment through 3D binaural sounds and haptic cues.

5. The audio games described in chapter 6, the Game Editor and the Binaural Game Analyzer tool

The audio games developed for this research employ 3D binaural sounds, auditory icons and earcons to improve user interaction. The Game Editor allows the experimenter to design new levels of the game, while the Binaural Game Analyzer tool offers statistical and real-time visualization of the players' performance.

6. **The multimodal training experiment for improving the gameplay performances of the visually impaired people**

This experiment was conducted in Györ, Hungary, with 10 visually impaired people (of which 2 were totally blind). The tools used for this study were the audio game described in 6.2., the Game Editor, the Game Analyzer and the haptic belt presented in 5.1.2., with the difference that all the 24 vibration motors have been placed on the headband, at 15 degrees difference all around the listener's head.

The multimodal (auditory and haptic) training helped the visually impaired subjects to adapt to altered hearing conditions (such as the use of 3D sounds filtered with non-individualized HRTFs) by creating a strong association between the auditory stimuli perceived in the headphones and the vibration corresponding to the direction of the sound source on the haptic belt. The subjects enhanced their spatial auditory resolution, navigational skills, orientation and mobility and decision making abilities. Moreover, they succeeded to build a solid spatial representation of the virtual environment that conducted to significant improvements in the game performance. The present 3D audio game can be considered an efficient modality for training the auditory perception and orientation skills of the users of a sensory substitution device. In addition to this, it is a useful rehabilitation tool for improving the navigational abilities of the visually impaired people in a safe and reliable environment, such as the virtual one. Nonetheless, the game can be played by the sighted people who want to try an alternative to the video games, entertain and train their sound localization skills at the same time.

7. The combination of white and pink noise in varying proportions, according to the direction of the sound source in space enabled both the sighted and the visually impaired subjects to achieve a higher sound localization accuracy, reflected in smaller angular precision errors and in a low incidence of front-back confusions. The results of our experiment are better than those obtained by Blum et al [38], Parseihian & Katz [42], Zahorik [55], Wenzel [127] and Padersen & Jorgensen [128].

In this research, we found evidence that the human auditory system is able to adapt to altered hearing conditions after a short training session based on multimodal (auditory, visual and haptic) feedback. The results of our test and training sessions proved that the sighted and visually impaired subjects have been able to use 3D binaural sounds synthesized from non-individualized HRTFs as the only means for navigating in a virtual auditory environment. Besides, the results of the post-test session demonstrate that the subjects enhanced their orientation and mobility skills, improved their directional decision-making abilities and recalibrated the spatial resolution when navigating in a virtual auditory environment and listening to both narrowband and broadband stimuli. In conclusion, the proposed approach can be considered a useful training and rehabilitation tool for the design of an assistive device for the blind people and for the future development of audio-only games.

7.2. FUTURE WORK

Future work involves the development of 3D navigational audio games for the visually impaired people on PC and mobile platforms, using the Csound library for Android [132]. Furthermore, the 3D sounds will integrate directional information in the median (vertical) plane, as well as a larger diversity of earcons, auditory icons and sound effects. These games will be part of the training procedure aimed to help the visually impaired people to safely and effectively use the sensory substitution device that will be developed by the Sound of Vision project [131].

ACRONYMS

HRTF - Head Related Transfer Function

SRT - Simple Reaction Time

MAA - Minimum Audible Angle

ILD - Interaural Level Difference

ITD - Interaural Time Difference

A1 - Primary Auditory Cortex

TTS - Text-to-Speech

SITREC - Stockholm International Toy Research Center

PRS - Peripheral Reflex System

CPU - Central Processing Unit

UWGD - USB Wireless Gateway Device

HAD - Haptic Actuator Device

ERM - Eccentric Rotating Mass

BIBLIOGRAPHY

1. World Health Organization Statistics. Available at: http://www.who.int/mediacentre/factsheets/fs282/en/. Accessed in June 2015.

2. *S. Spagnol, M. Geronazzo, D. Rocchesso, F. Avanzini*, "Extraction of Pinna Features for Customized Binaural Audio Delivery on Mobile Devices", Proceedings of International Conference on Advances in Mobile Computing & Multimedia, Vienna, Austria, 2-4 December 2013.

3. *M. Dellepiane, N. Pietroni, N. Tsingos, M. Asselot, R. Scopigno*, "Reconstructing Head Models from Photographs for Individualized 3D-audio Processing", Comput. Graph. Forum, **vol**. 27, no. 7, 2008, pp. 1719-1727.

4. *A. Meshram, R. Mehra, H. Yang, E. Dunn, J.M. Frahm, D. Manocha*, "P-HRTF: Efficient Personalized HRTF Computation for High-Fidelity Spatial Sound", International Symposium on Mixed and Augmented Reality (ISMAR), Munich, Germany, 10-12 September 2014.

5. *D.N. Zotkin, J. Hwang, R. Duraiswami, L.S. Davis*, "HRTF Personalization Using Anthropomentric Measurements", Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA'03), 2003, pp. 157-160.

6. *J. Ahveninen, N. Kopčo, I.P. Jääskeläinen*, "Psychophysics and neuronal bases of sound localization in humans", Hearing Research, **vol.** 307, 2014, pp. 86-97.

7. *O. Bălan, A. Moldoveanu, F. Moldoveanu*, "A Systematic Review of the Methods and Experiments Aimed to Reduce Front-Back Confusions in the Free-Field and Virtual Auditory Environments", International Journal of Acoustics and Vibration (in review)

8. *O. Bălan, A. Moldoveanu, A. Butean, F. Moldoveanu, I. Negoi*, "Comparative Research on Sound Localization Accuracy in the Free-Field and Virtual Auditory Displays", Proceedings of The 11th International Scientific Conference eLearning and software for Education, Bucharest, April 23-24, 2015.

9. *O. Bălan, A. Moldoveanu, F. Moldoveanu, I. Negoi*, "The Role of Perceptual Feedback Training on Sound Localization Accuracy in Audio Experiments", Proceedings of The 11th International Scientific Conference eLearning and software for Education, Bucharest, April 23-24, 2015

10. *Ó.I. Jóhannesson, O. Bălan, R. Unnthorsson, A. Moldoveanu, Á. Kristjánsson*, "The Sound of Vision: On the Feasibility of Creating a Representation of the Auditory Environment for Spatial Navigation for the Visually Impaired", Frontiers in Human Neuroscience (in review)

11. *C.J. Plack*, The Sense of Hearing, Psychology Press, 2005.

12. *T.R. Letowski, S.T. Letowski*, Auditory Spatial Perception: Auditory Localization, Army Research Laboratories ARL-TR-6016, 2012.

13. Sound loudness. Available at: https://en.wikipedia.org/wiki/Loudness. Accessed in June 2015.

14. *S. Carlile*, Psychoacoustics, In Thomas Hermann, Andy Hunt, John G. Neuhoff (Eds.), The Sonification Handbook, (pp. 41-61). Berlin, Germany: Logos Verlag Berlin, 2011.

15. Sound pitch. Available at: https://en.wikipedia.org/wiki/Pitch. Accessed in June 2015.

16. *T. Letowski, S. Letowski*, Localization Error: Accuracy and Precision of Auditory Localization (Chapter 4), Advances in Sound Localization, book edited by Pawel Strumillo, Intech, 2011.

17. *H. Wallach*, "The Role of Head Movements and Vestibular and Visual Cues in Sound Localization", Journal of Experimental Psychology, **vol**. 27, no. 4, 1940, pp. 339-368.

18. *A. Kudo, H. Higuchi, H. Hokari, S. Shimada*, "Improved Method for Accurate Sound Localization", Acoustical Science and Technology, **vol**. 27, no. 3, 2006, pp. 134-146.

19. The Synthetic Psychology of Sound Localization. Available at: http://gureckislab.org/courses/spring13/robots/SoundLocalization-2.html. Accessed in June 2015.

20. Spatial Hearing. Available at: http://music.miami.edu/programs/mue/Research/jwest/Chap_2/Chap_2_Spatial_Hearing.html. Accessed in June 2015.
21. *B.G. Shinn-Cunningham, S. Santarelli, N. Kopco*, "Tori of Confusion: Binaural Localization Cues for Sources Within Reach of a Listener", J. Acoust. Soc. Am., **vol**. 107, no. 3, 2000, pp. 1627-1636.
22. Head Related Transfer Functions. Available at: http://interface.cipic.ucdavis.edu/sound/tutorial/hrtf.html. Accessed in June 2015.
23. Introducing the HRTF. Available at: http://alumnus.caltech.edu/~franko/thesis/Chapter4.html. Accessed in June 2015.
24. *M. Kin, P. Plaskota*, "Comparison of Sound Attributes of Multichannel and Mixed-Down Stereo Recordings", Archives of Acoustics, **vol**. 36, no. 2, 2011, pp. 333-345
25. *Z. Xiao-li, X. Bo-sun Xie*, Chapter 6 - Head-Related Transfer Functions and Virtual Auditory Display, Engineering "Electrical and Electronic Engineering: "Soundscape Semiotics - Localization and Categorization", book edited by Herve Glotin, 2014
26. *K.I McAnally, R.L. Martin*, "Sound Localization with Head Movement: Implications for 3D Audio Display", Frontiers in Neuroscience, **vol**. 8, no. 2010, 2014, pp. 1-6
27. *F. Wightman*, "Resolution of Front-Back Ambiguity in Spatial Hearing by Listener and Source Movement", J. Acoust. Soc. Am., **vol**. 105, no. 5, 1999, pp. 2841-2853.
28. *D.R. Begault, E.M. Wenzel, M.R. Anderson*, "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., **vol**. 49, no. 10, 2001, pp. 904-916.
29. *P.A. Hill, P.A. Nelson, O. Kirkeby*, "Resolution of Front-Back Confusion in Virtual Acoustic Imaging Systems", The Journal of the Acoustical Society of America, **vol**. 108, no. 6, 2000, pp. 2901-2910.
30. *N. Mariette*, "Mitigation of Binaural Front-Back Confusions by Body Motion in Audio Augmented Reality". Proceedings of the 13th International Conference on Auditory Display, Montreal, Canada, June 26-29, 2007.
31. *A. Nykanen, A. Zedigh, P. Mohlin*, "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.
32. *J. Towers, R. Burgess-Limerick, S. Riek*, "Improving 3-D Audio Localisation Through the Provision of Supplementary Spatial Audio Cues", The Open Ergonomics Journal, **vol**. 5, no. , 2013, pp. 1-9.
33. *D.K. Reed, R.C. Maher*, "An Investigation of Early Reflection's Effect on Front-Back Localization in Spatial Audio", Proc. AES 127th Convention, New York, NY, USA, 2009
34. *D.L. Wang, G.J. Brown*, Computational Auditory Scene Analysis: Principles, Algorithms, and Applications, Wiley-IEEE Press, 2006
35. *P.M. Hofman, J.G. Van Riswick, A.J. Van Opstal*, "Relearning Sound Localization with New Ears", Nat Neurosci, **vol.** 1, no. 5, September 1998, pp. 417-421.
36. *S. Carlile, P. Leong, S. Hyams*, "The Nature and Distribution of Errors in Sound Localization by Human Listeners", Hear. Res., **vol**. 114, no. 1-2, December 1997, pp. 179-196.
37. *J.C. Makous, J.C. Middlebrooks*, "Two-Dimensional Sound Localization by Human Listeners", J. Acoust. Soc. Am., **vol**. 87, no. 5, May 1990, pp. 2188-2200.
38. *A. Blum, B.F.G. Katz, O. Warusfel*, "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), Strasboug, France. 2004. pp. 1225–1226.

39. *A. Honda, H. Shibata, S. Hidaka, J. Gyoba, Y. Iwaya, Y. Suzuki*, "Effects of Head Movement and Proprioceptive Feedback in Training of Sound Localization", Iperception, **vol**. 4, no. 4, June 2013, pp. 253-264 doi: 10.1068/i0522.
40. *R.L. Goldstone*, "Perceptual Learning", Annual Review of Psychology, **vol.** 49, 1998, pp.585-612
41. *C. Mendonça*, "A Review on Auditory Space Adaptations to Altered Head-Related Cues", Front. Neurosci., **vol**. 8, no. 219, 2014.
42. *G. Parseihian, B.F.G. Katz*, "Rapid Head-Related Transfer Function Adaptation Using a Virtual Auditory Environment", J. Acoust. Soc. Am., **vol.** 131, 2012, pp. 2948-2957. doi: 10.1121/1.3687448.
43. *R.W. Bauer, J.L. Matuzsa, R.F. Blackmer, S. Glucksberg*, "Noise Localization after Unilateral Attenuation", J. Acoust.Soc.Am., **vol**. 40, pp. 441–444.doi: 10.1121/1.1910093
44. *M. Florentine*, "Relation Between Lateralization and Loudness in Asymmetrical Hearing Losses", J. Am.Audiol.Soc., **vol**. 1, 1976, pp. 243–251.
45. *A.D. Musicant, R.A. Butler*, "Monaural Localization: An Analysis of Practice Effects", Percept.Psychophys, **vol**. 28, 1980, pp.236–240.doi:10.3758/BF03 204379.
46. *W.H. Slattery, J.C. Middlebrooks*, "Monaural Sound Localization: Acute verses Chronic Unilateral Impairment", Hear.Res., **vol**. 75, 1994, pp. 38–46.
47. *D.P. Kumpik, O. Kacelnik, A.J. King*, "Adaptive Reweighting of Auditory Localization Cues in Response to Chronic Unilateral Ear Plugging in Humans", J. Neurosci., **vol.** 30, pp. 4883–4894. doi:10.1523/jneurosci.5488-09.2010.
48. *M.M. Van Wanrooij, A.J. Van Opstal*, "Relearning Sound Localization with a New Ear", J. Neurosci., **vol.** 25, 2005, pp. 5413–5424.doi:10.1523/jneurosci.0850-05.2005
49. *R. Held*, "Shifts in Binaural Localization After Prolonged Exposures to a Typical Combinations of Stimuli", The American Journal of Psychology, **vol.** 68, no. 4, December 1955, pp. 526-548.
50. *S. Carlile*, "The Plastic Ear and Perceptual Relearning in Auditory Spatial Perception", Frontiers in Neuroscience, **vol**. 8, 2014. doi: 10.3389/fnins.2014.00237.
51. *E.I. Knudsen*, "Mechanisms of Experience-Dependent Plasticity in the Auditory Localization Pathway of the Barn Owl", J. Comp. Physiol., **vol**. 185, 1999, pp. 305–321.
52. *B.A. Linkenhoker, C.G. von der Ohe, E.I. Knudsen*, "Anatomical Traces of Juvenile Learning in the Auditory System of Adult Barn Owls", Nat. Neurosci., **vol**. 8, 2005, pp. 93–98.
53. *S. Carlile, T. Blackman*, "Relearning Auditory Spectral Cues for Locations Inside and Outside the Visual Field", J. Assoc.Res.Otolaryngol. **vol.** 15, 2013, pp. 249–263. doi: 10.1007/s10162-013-0429-5.
54. *R.S. Woodworth, H. Schlosberg*, Experimental Psychology, Holt, Rinehart and Winston, New York, 1954
55. *P. Zahorik, P. Bangayan, V. Sundareswaran, K. Wang, C. Tam*, "Perceptual Recalibration in Human Sound Localization: Learning to Remediate Front-Back Reversals", J. Acoust. Soc. Am., **vol.** 120, 2006, pp. 343-359. doi: 10.1121/1.2208429
56. *C. Mendonça, G. Campos, P. Dias, J.A. Santos*, "Learning Auditory Space: Generalization and Long-Term Effects", PLoS ONE, **vol**. 8, no. 3, 2013. doi:10.1371/journal.pone.0077900
57. *V. Tabry, R.J. Zatorre, P. Voss*, "The Influence of Vision on Sound Localization Abilities in Both the Horizontal and Vertical Planes", Front Psychol., **vol**. 4, 932, 2013.
58. *O. Bălan, A. Moldoveanu, F. Moldoveanu*, "Spatial Auditory Representation in the Case of the Visually Impaired People", Journal of Information Systems & Operations Management, **vol.** 9, no. 1, May 2015, pp. 1-11.
59. O. *Bălan, A. Moldoveanu, F. Moldoveanu, M. I. Dascălu*, "Navigational 3D Audio-Based Game - Training Towards Rich Auditory Spatial Representation of the Environment", Proceedings of the 18th

International Conference on System Theory, Control and Computing, Sinaia, Romania, October 17-19, 2014, pp. 688-693.
60. *P. Majdak, M.J. Goupell M, B. Laback*, "3-D Localization of Virtual Sound Sources: Effects of Visual Environment, Pointing Method, and Training", Atten Percept Psychophys, **vol.** 72, no. 2, Feb 2010, pp. 454–469.
61. *B. Jones, B. Kabanoff*, "Eye Movements in Auditory Space Perception", Percept. Psychophys, **vol.** 17, 1975, pp. 241–245.
62. *P. Zahorik, C. Tam, K. Wang, P. Bangayan, V. Sundareswaran*, "Localization Accuracy in 3D Sound Displays: The role of Visual-Feedback Training", in P. N. Rose (Ed.), Proceedings of the Advanced Displays Consortium: ARL's 5th Federated Laboratory Annual Symposium, 2001, College Park, MD. pp. 17-22.
63. *Q. Chen, M. Zhang, X. Zhou*, "Spatial and Nonspatial Peripheral Auditory Processing in Congenitally Blind People", Neuroreport., **vol**. 17, no. 13, September 2006, pp. 1449-1452.
64. *O. Bălan, A. Moldoveanu, F. Moldoveanu, A. Morar*, "Assistive IT for Visually Impaired People", Journal of Information Systems & Operations Management, **vol.** 7 no.2, 2013, pp. 391-404.
65. *B. Röder, W. Teder-Sälejärvi, A. Sterr, F. Rösler, S.A. Hillyard, N.J. Neville*, "Improved Auditory Spatial Tuning in Blind Humans", Nature, **vol**. 400(6740), July 1999, pp. 162-166.
66. *F. Gougoux, R.J. Zatorre, M. Lassonde, P. Voss, F. Lepore*, "A Functional Neuroimaging Study of Sound Localization: Visual Cortex Activity Predicts Performance in Early-Blind Individuals", PLoS Biol., **vol**. 3, no. 2, February 2005.
67. *J.P. Rauschecker, M. Kort*, "Auditory Compensation for Early Blindness in Cat Cerebral Cortex", J Neurosci., **vol**. 13, 1993, pp. 4538–4548.
68. *T. Korte, J.P. Rauschecker*, "Auditory Spatial Tuning of Cortical Neurons is Sharpened in Cats with Early Blindness", J Neurophys., **vol**. 70, 1993, pp. 1717–1721.
69. *P. Voss, M. Lassonde, F. Gougoux, J.P. Guillemot, F. Lepore*, "Early - and Late-Onset Blind Individuals Show Supra-Normal Auditory Abilities in Far Space", Curr. Biol., **vol.** 14, 2004, pp. 1-5.
70. *O. Collignon, G. Vandewalle, P. Voss, G. Albouy, G. Charboneau, M. Lassonde, F. Lepore*, "Functional Specialization for Auditory–Spatial Processing in the Occipital Cortex of Congenitally Blind Humans", Proc. Natl. Acad. Sci. U. S. A., **vol**. 108, 2011, pp. 4435–4440.
71. *C.H. Chan, W.K.A. Wong, K-H Ting, S. Whitfield-Gabrieli, J. He, M.C.T. Lee*, "Cross Auditory-Spatial Learning in Early-Blind Individuals", Human Brain Mapping, 2011.
72. *T. Elbert, A. Sterr, B. Rockstroh, C. Pantev, M.M. Müller*, "Expansion of the Tonotopic Area in the Auditory Cortex of the Blind", J. Neurosci., **vol.** 2, 2002, pp.9941–9944.
73. *S. Maidenbaum, S. Abboud, A. Amedi*, "Sensory Substitution: Closing the Gap Between Basic Research and Widespread Practical Visual Rehabilitation", Neuroscience and Biobehavioral Reviews, 2014, http://dx.doi.org/10.1016/j.neubiorev.2013.11.007. Published online November 23, 2013
74. *N. Lessard, M. Pare, F. Lepore, M. Lassonde*, "Early-Blind Human Subjects Localize Sound Sources Better than Sighted Subjects", Nature, 1998, **vol**. 395, pp. 278–280
75. *M.E. Doucet, J.P. Guillemot, M. Lassonde, J.P. Gagné, C. Leclerc, F. Lepore*, "Blind Subjects Process Auditory Spectral Cues More Efficiently than Sighted Individuals", Exp. Brain. Res., **vol.** 160, no. 2, January 2005, pp. 194-202.
76. *M. Ohuchi, Y. Iwaya, Y. Suzuki, T. Munekata*, "A Comparative Study of Sound Localization Acuity of Congenital Blind and Sighted people", Acoustical Science and Technology, **vol.** 27, 2006, pp. 290-293.
77. *Gy. Wersényi*, "Virtual Localization by Blind Persons", JAES, **vol**. 60, no. 7/8, July 2012, pp. 568-579.

78. *M.P. Zwiers, A.J. Van Opstal, J.R. Cruysberg*, "Two-Dimensional Sound-Localization Behavior of Early-Blind Humans", Exp Brain Res., **vol.** 140, no. 2, September 2001, pp. 206-222.
79. *J. Gonçalves da Silva, A.G. Pascoal, M.L.M. Pereira*, "Spatial Localization of Sounds in Blind Individuals: Comparative Study of Three Dimensional Position of the Head in Blind and Non Blind Adults", Rev. bras. educ. espec. [online]., vol.14, no.1, 2008, pp. 111-120.
80. *M. Bujacz, P. Skulimowski, P. Strumiłło*, "Sonification of 3D Scenes Using Personalized Spatial Audio to Aid Visually Impaired Persons", The 17th International Conference on Auditory Display (ICAD-2011) June 20-24, 2011, Budapest, Hungary.
81. *B.F.G. Katz, L. Picinali*, "Spatial Audio Applied to Research with the Blind", Advances in Sound Localization, InTech, 2011. Available at: http://www.intechopen.com/books/advances-in-sound-localization/spatial-audio-applied-to-research-with-theblind.
82. *O. Bălan, A. Moldoveanu, F. Moldoveanu*, "Navigational Audio Games: An Effective Approach Towards Improving Spatial Contextual Learning for Blind People", The International Journal on Disability and Human Development, vol. 14, no. 2, April 2015, pp. 109-118, ISSN (Online) 2191-0367, ISSN (Print) 2191-1231, DOI: 10.1515/ijdhd-2014-0018
83. *O. Bălan, A. Moldoveanu, F. Moldoveanu, O.M. Ferche*, "O Scurtă Clasificare a Jocurilor Audio în Contextul Îmbunătăţirii Interacţiunii şi Accesibilităţii pentru Persoanele Nevăzătoare", Revista Română de Interacţiune Om-Calculator, vol. 7, no. 3, 2014, pp. 225-238, ISSN 1843-4460
84. *O. Bălan, A. Moldoveanu, F. Moldoveanu*, ICT for Visually Impaired, Printech, ISBN 978-606-23-0348-8, 2014.
85. *B. Yuan*, "Towards Generalized Accessibility of Video Games for the Visually Impaired", PhD Thesis, 2009. Available at : http://dl.acm.org/citation.cfm?id=1713911
86. *E.C. Connors, R.E. Chrastil, J. Sanchez, L.B. Merabet*, "Action Video Game Play and Transfer of Navigation and Spatial Cognition Skills in Adolescents Who Are Blind", Front Hum Neurosci., **vol.** 11, 2014. doi: 10.3389/fnhum.
87. *D. Archambault, R. Ossmann, T. Gaudy, K. Miesenberger*, "Computer Games and Visually Impaired People", Digital Journal of CEPIS, 2007, pp. 43-53
88. *J. Chen*, "Flow in Games (and Everything Else)", Communications of the ACM, **vol**. 50, no. 4, 2007, pp. 31-34.
89. *M. Csíkszentmihályi*, Flow: The Psychology of Optimal Experience, Harper and Row, New York, 1990
90. *T. Gaudy, S. Natkin, D. Archambault*, "Pyvox 2: An Audio Game Accessible To Visually Impaired People Playable Without Visual Nor Verbal Instructions", Transaction on Edutainment II LNCS 5660, 2009, pp. 176-186
91. *J.R. Parker, J. Heerema*, "Audio Interaction in Computer Mediated Games", International Journal of Computer Games Technology, **vol.** 2008, Article ID 178923.
92. *Á Csapó, Gy. Wersényi*, "Overview of Auditory Representations in Human-Machine Interfaces", Journal ACM Computing Surveys (CSUR), **vol.** 46, no. 2, 2013.
93. *K. Jorgensen*, "On the Functional Aspects of Computer Game Audio", Proceedings of the Audio Mostly Conference, 2006
94. *N. Rober, M. Masuch*, "Playing Audio-Only Games- A Compendium of Interactions with Virtual, Auditory Worlds", Proceedings of DiGRA Conference, 2005
95. *J. Carvalho, T. Guerreiro, L. Duarte, L. Carrico*, "Audio-Based Puzzle Gaming for Blind People", 2nd Workshop on Mobile Accessibility, 2012
96. *T. Stockman, N. Rajgor, O. Metatla, L. Harrar*, "The Design of Interactive Audio Soccer", Proceedings of the 2nd International Workshop on Interactive Sonification, 2007

97. *L. B. Merabet, J. Sanchez*, "Audio-Based Navigation Using Virtual Environments: Combining Technology and Neuroscience", AER Journal: Research and Practice in Visual Impairment and Blindness, **vol**. 2, no. 3, 2009, pp. 128-137.
98. Enhancing the Navigational Skills of the Blind: Creating a "Soundscape" Virtual Audio Environment. Available at: http://ctdinstitute.org/library/2014-10-06/enhancing-navigational-skills-blind-creating-%E2%80%9Csoundscape%E2%80%9D-virtual-audio. Accessed in June 2015.
99. *J. Friberg, D. Gardenfors*, "Audio Games: New Perspectives on Game Audio", In Proceedings of ACE 2004, Singapore, June 3-5, 2004
100. *T. Westin*, "Game Accessibility Case Study: Terraformers- a Real-Time 3D Graphics Game", Proc. of the The Fifth International Conference on Disability, Virtual Reality and Associated Technologies, Oxford, UK, 2004
101. *J. Sanchez, M. Lumbreras*, "Usability and Cognitive Impact of the Interaction with 3D Virtual Interactive Acoustic Environments by Blind Children", Proc. 3rd Intl Conf. Disability, Virtual Reality & Assoc. Tech., Alghero, Italy 2000
102. The Blind Side. Available at: http://michaeltastolfi.com/#/blindside/. Accessed in June 2015.
103. A Sound Solution: History of Audio Games for the Visually Impaired. Available at: http://artistryingames.com/sound-solution-history-audio-games-visually-impaired/. Accessed in June 2015.
104. *O. Bălan, F. Moldoveanu, A. Morar, V. Asavei*, "Experiments on Training the Sound Localization Abilities: A Systematic Review", eLSE Conference, Bucharest, 2014
105. *J. Sanchez, F. Aguayo*, "AudioGene: Mobile Learning Genetics through Audio by Blind Learners" Learning to Live in the Knowledge Society, 2008, pp. 79-86.
106. *J. Sanchez, H. Flores, M. Saenz*, "Mobile Science Learning for the Blind", Proceedings of the CHI Conference, Florence, Italy, 2008
107. *O. Bălan, A. Moldoveanu, H. Nagy, G. Wersényi, N. Botezatu, A. Stan, R. G. Lupu*, "Haptic-Auditory Perceptual Feedback Based Training for Improving the Spatial Acoustic Resolution of the Visually Impaired People", The 21st International Conference on Auditory Display (ICAD-2015), July 8-10, 2015, Graz, Austria
108. *O. Bălan, A. Moldoveanu, F. Moldoveanu*, "3D Audio and Haptic Interfaces for Training the Spatial Acoustic Resolution in Virtual Auditory Environments", The 21st International Conference on Auditory Display (ICAD-2015), July 8-10, 2015, Graz, Austria
109. *O. Bălan, A. Moldoveanu, F. Moldoveanu*, "Perceptual Feedback Training for Improving Spatial Acuity and Resolving Front-Back Confusion Errors in Virtual Auditory Environments", Archives of Acoustics (in review)
110. *O. Bălan, A. Moldoveanu, F. Moldoveanu*, "Comparative Analysis on How Multimodal Perceptual Feedback Training Improves the Spatial Auditory Performance of the Sighted and Visually Impaired People", Archives of Acoustics (in review)
111. *O. Bălan, A. Moldoveanu, F. Moldoveanu*, "Binaural Sound Analysis and Spatial Localization for the Visually Impaired People", 9th International Conference on Interfaces and Human Computer Interaction, 22 – 24 July 2015, Las Palmas de Gran Canaria, Spain
112. *A. Stan, N. Botezatu, R.G. Lupu*, "The Design of a Scalable Haptic System Used for Impaired People Assistance", CSCS20: The 20th International Conference on Control Systems and Computer Science, 27-29 May 2015, Bucharest, Romania
113. Touch Sense Library. Available at: http://www.ti.com/tool/drv2605ldgsevm-m. Accessed in June 2015.
114. *A. Stan, N. Botezatu*, "Introducing aceMote: An energy efficient 32 bit mote", International Conference on Development and Application Systems (DAS), pp. 79-84, Suceava, Romania, 2014

115. Precision Microdrives, Eccentric Rotating Mass (ERM) Vibration. Available at: http://www.precisionmicrodrives.com/application-notes-technical-guides/application-bulletins/ab-004-understanding-erm-characteristics-for-vibration-applications. Accessed in June 2015.
116. MCP2200. Available at: http://www.microchip.com/wwwproducts/devices.aspx?dDocName=en546923#documentation. Accessed in June 2015.
117. MIT HRTF Database. Available at: http://sound.media.mit.edu/resources/KEMAR.html. Accessed in June 2015.
118. Csound manual. Available at: http://floss.booktype.pro/csound/preface/. Accessed in June 2015.
119. *V. Lazzarini, B. Carty*, "New Csound Opcodes for Binaural Processing", In Proc. 6th Linux Audio Conference, Cologne, Germany, 2008, pp. 28-35
120. Csound QT. Available at: http://qutecsound.sourceforge.net/. Accessed in August 2015.
121. Visual Studio. Available at: https://www.visualstudio.com/. Accessed in August 2015.
122. The C# programming language. Available at: https://msdn.microsoft.com/en-us/library/67ef8sbd.aspx. Accessed in August 2015.
123. Inverse square law. Available at: http://hyperphysics.phy-astr.gsu.edu/hbase/forces/isq.html. Accessed in August 2015.
124. ANOVA statistic test. Available at: https://en.wikipedia.org/wiki/Analysis_of_variance. Accessed in August 2015.
125. Student t-test. Available at: https://en.wikipedia.org/wiki/Student%27s_t-test. Accessed in August 2015.
126. Wilcoxon statistic test. Available at: https://en.wikipedia.org/wiki/Wilcoxon_signed-rank_test. Accessed in August 2015.
127. *E.M. Wenzel*, "Effect of Increasing System Latency on Localization of Virtual Sounds with Short and Long Duration", Proceedings of the International Conference on Auditory Display, Espoo, Finland, July 29-August 1, 2001, 185–190.
128. *J.A. Padersen, T. Jorgensen*, "Localization Performance of Real and Virtual Sound Sources", Proceedings North Atlantic Treaty Organization, 2005
129. *O. Bălan, A. Moldoveanu, F. Moldoveanu, G. Wersényi*, Improving the Audio Game Playing Performances of the Visually Impaired People through Multimodal Training, Journal of Visual Impairment & Blindness (in review)
130. *O. Bălan, A. Moldoveanu, F. Moldoveanu, A. Butean*, "Developing a Navigational 3D Audio Game with Hierarchical Levels of Difficulty for the Visually Impaired Players", a 12-a Conferinţă Naţională de Interacţiune Om-Calculator, 24-25 September 2015, Bucharest, Romania
131. The Sound of Vision website. Available at: http://www.soundofvision.net/. Accessed in August 2015.
132. The Csound Library for Android. Available at: http://sourceforge.net/projects/csound/files/csound5/Android/. Accessed in August 2015.
133. *T. Iachini, G. Ruggiero, F. Ruotolo*, "Does Blindness Affect Egocentric and Allocentric Frames of Reference in Small and Large Scale Space", Behav Brain Res, 273, 2014, pp. 73–81.
134. *M. Eid, A. El Issawi, A. El Saddik*, "Slingshot 3D: A Synchronous Haptic-Audio-Video Game", Multimed Tools Appl, **vol**. 71, 2014, pp. 1635-1648
135. *J. Sanchez, M. Campos, M. Espinoza, L.B. Merabet*, "Audio Haptic Videogaming for Developing Wayfinding Skills in Learners who are Blind", Proceedings of the 19th international conference on Intelligent User Interfaces, 2014
136.*K. Strelnikov, M. Rosito, P. Barone*, "Effect of Audiovisual Training on Monaural Spatial Hearing in Horizontal Plane", PLoS ONE, **vol**. 6, no. 3, 2011

APPENDICES

APPENDIX 1

The “whitepink.csd” file that generates the white-pink noise 3D sound in CSound QT:

```
<CsOptions>
-o dac
</CsOptions>
<CsInstruments>
sr = 44100
ksmps = 128
nchnls = 2
0dbfs = 1.0
#define FILE1 #"hrtf-44100-left.dat"#
#define FILE2 #"hrtf-44100-right.dat"#

instr 1
kAzimuth chnget "azimuth1"
kElevation = 0   ;chnget "elevation"
kAmp chnget "ampli1"
kAmpVar = 0.3
kSinLFO    oscili kAmpVar, 1/2
kSquareLFO oscili kAmpVar, 1/4, 3
kValue chnget "kvalw1"
aSignal noise  kAmp * (1+kSinLFO*(kAmpVar+kSquareLFO)), 1
aleft, aright  hrtfmove2  aSignal, kAzimuth, kElevation, $FILE1, $FILE2 ;[,ioverlap, iradius, isr]
outs aleft*kValue, aright*kValue
endin

instr 2
kAzimuth chnget "azimuth1"
kElevation = 0   ;chnget "elevation"
kValue  chnget "kvalp1"
kAmp chnget "ampli1"
awhite unirand 2.0
white = awhite - 1.0
apink  pinkish kAmp*awhite, 1, 0, 0, 1
aleft, aright  hrtfmove2  apink, kAzimuth, kElevation, $FILE1, $FILE2 ;[,ioverlap, iradius, isr]
outs aleft*kValue, aright*kValue
endin
</CsInstruments>
<CsScore>
f 1 0 1024 10 1 ; sine
f3 0 16384 10 1 0   0.3 0    0.2 0     0.14 0     .111   ; Square
i 1 0 3600
```

```
i 2 0 3600
e

</CsScore>
</CsoundSynthesizer>
<bsbPresets>
</bsbPresets>
<bsbPanel>
 <label>Widgets</label>
<objectName/>
<x>100</x>
<y>100</y>\
<width>320</width>
<height>240</height>
 <visible>true</visible>
 <uuid/>
<bgcolor mode="nobackground">
 <r>255</r>
<g>255</g>
 <b>255</b>
</bgcolor>
</bsbPanel>
<bsbPresets>
</bsbPresets>
```

APPENDIX 2

The “ding.csd” file that generates the “ding” 3D sound in CSound QT:

```
<CsOptions>
-o dac
</CsOptions>
<CsInstruments>
sr = 44100
ksmps = 128
nchnls = 2
0dbfs = 1.0
#define FILE1 #"hrtf-44100-left.dat"#
#define FILE2 #"hrtf-44100-right.dat"#

instr 1
kAzimuth chnget "azimuth1"
kElevation = 0   ;chnget "elevation"
kAmp chnget "ampli1"
kAmpVar    = 0.1 ; amplitude variation to introduce iregularities in the noise
printk2   kAzimuth
kSinLFO    oscili kAmpVar,  1/2
kSquareLFO oscili kAmpVar, 1/4, 3
```

```
kBeta chnget "filter1"
ain diskin "ding.wav",1,1,1
ain2=ain*kAmp
aleft, aright  hrtfmove2  ain2, kAzimuth, kElevation, $FILE1, $FILE2 ;[,ioverlap, iradius, isr]
outs aleft, aright
endin

</CsInstruments>
<CsScore>
 i 1 0 3600
  e
</CsScore>
</CsoundSynthesizer>
<bsbPresets>
</bsbPresets>
<bsbPanel>
<label>Widgets</label>
<objectName/>
<x>100</x>
<y>100</y>
<width>320</width>
<height>240</height>
<visible>true</visible>
<uuid/>
 <bgcolor mode="nobackground">
<r>255</r>
<g>255</g>
<b>255</b>
</bgcolor>
</bsbPanel>
<bsbPresets>
</bsbPresets>
```

APPENDIX 3

The format of the declarations completed and signed by the subjects after the sound localization experiments performed in Györ, Hungary:

DECLARATION

I, ____, born on ______, living in ___________________________________suffering from (the name of the visual impairment ____ percent of vision ____________________________I have participated at the haptic-auditory perceptual localization experiments that took place on 2.02.2015, 5.02.2015, 9.02.2015 and 12.02.2015 in Gyor, Hungary.

Date, Signature

DECLARATION

I, ____, born on______________________, living in__________________, suffering from (the name of the visual impairment)____, percent of vision _________________.

I have participated at the haptic-auditory perceptual localization experiments (audio game playing and training sessions) that took place between 20.04.2015 and 24.04.2015 in Gyor, Hungary.

Date, Signature