PhD THESIS

Design, modeling and analysis of object localization through acoustical signals for cognitive electronic travel aid for blind people

Department of Graphics Engineering

School of Design Engineering

Author:

Larisa Dunai

(Electronic engineer)

Thesis director:

Guillermo Peris Fajarnes

Beatriz Defez García

UNIVERSIDAD POLITECNICA DE VALENCIA

Design, modeling and analysis of object localization through acoustical signals for cognitive electronic travel aid for blind people

Department of Graphics Engineering

School of Design Engineering

Author:
Larisa Dunai
(Electronic engineer)

Thesis director:
Guillermo Peris Fajarnes
Beatriz Defez García

July 2010

Acknowledgments

Así como los ojos de los murciélagos
se ofuscan a la luz del día,
de la misma manera a la inteligencia
de nuestra alma la ofuscan las cosas evidentes

Aristóteles

This thesis and, especially, the acknowledgments represent the end of a quite long journey through the field of the research spent roaming around different countries. Such a journey would not have been possible without the support and encouragement of my family and especially to my husband Jose Alfonso and my mother Elena, parents-in-law Francisca and Jose Alfonso and brothers.

Also, and not at the final, to my thesis director, Guillermo Peris Fajarnes, director of the Research Center in Graphic Technology and Beatriz Defez Garcia for their support, without which the realization of this hard work would have been unable.

I would also like to thank the staff of Instituto dei Cehi Cavatza from Italy for letting me absorb through countless conversations their knowledge on blind people as well as their necessities. I thank Giovanni Celucci and Giovanni Ciafonni for the help with the tests with blind people.

I also thank Hans Kalvasser from Deutscher Blinden- und Sehbehindertenverband e. V. for its detailed reports regarding blind users necessities and comments during the trials.

I would like to thank Antonio Rodriguez and Jose Luis Gonzales Mora and their group at Universidad de la Laguna for their help in the area of psychoacoustics and the development of the acoustical system and HRTFs generation.

Finally, I thank Majid Mirmehdi and his group at University of Bristol for their great work in the development of the object detection algorithm.

Abstract

Blind and impaired people need for an assistance device able to provide them independent mobility. The design and development of such a navigation device will mean a significant advance in engineering and research. In these previous decades, many researchers have been investigating on different methods for environment information representation, able to be implemented in electronic travel aids for blind and impaired people.

The present thesis carries out the design, modelling, implementation, experimentation and analysis of a wearable object detector and navigation device for blind people, named Cognitive Aid System for Blind People, CASBliP. The CASBliP device represents an Electronic Travel Aid, whose primary goal is to help blind users to navigate independently and safety both in indoor and outdoor environments.

In this context, the thesis begins with a detailed presentation of the state-of-the-art of the nowadays existing electronic travel aid systems for blind people, as well as those under development. This review comprises devices developed from the Second World War, when the development of sensors played an important role in the human life, until nowadays. In this initial chapter, a classification of the electronic travel aid systems into three main groups, based on the type of system, is presented: obstacle detectors, environmental sensors and navigation systems. More than forty relevant systems are described, explaining the main differences between the different categories. Despite the increasing knowledge and wide usage of the sensory electronic travel aid systems in the world, the development of a universal and more accurate navigation and object detection system has not been achieved yet.

In order to achieve the thesis objective, the designed device consists of two inputs, one output, a portable computer and a FPGA as processing units, which can function individually. The input system consist of an array of 64×1 CMOS Time of Flight sensors attached to a pair of glasses and two cameras mounted on a helmet. The system output consists of a pair of mini stereo headphones through which the user will perceive the environmental objects and free paths. The input system goal is to capture environmental information from the user's direction of view. Taking as a basis the 3D environmental information perceived, the input system represents the moving objects and detects all static and moving objects and free paths using the depth maps, segmentations algorithms and motion detection algorithms. The high resolution input of the image is projected onto high resolution acoustic sounds

via a methodology based on navigation criteria and models based on convolution with non-individual Head-Related Transfer Functions.

The device implements a methodology for simulating that a series of sounds, in a virtual way, are radiated by the user surrounding objects, sounds which are capable to carry accurate spatial information. The idea is to generate in the user a correct perception of virtual sound sources emitting from the object surfaces, which aims to allow the brain to create a three-dimensional perceptual image of those objects like in the real world. Using these ideas, it intends to create a global perception of the sound, enabling blind people, on real-time, to perceive and get a global image of the surrounding environment and the way the objects are organized.

It is well known that humans use a wide range of information for navigation as vision, feelings and hearing. When human visual system is damaged (loss of vision), the hearing system gets the main role on navigation. It is extremely important to analyze and define the aspects of the visual scene, which represent the most important features for navigation and object identification, in order to represent the object presence and determine its position in space.

In the third chapter, an overview of the auditory system is described including its basic components and auditory organisation. This chapter provides a background of sound localization with acoustic cues (monaural and binaural cues, interaural time difference and interaural level difference, reverberation effect, cone of confusion, precedence effect and cross-correlation model). This review is aimed to represent the whole picture of the level of accuracy for sound localization. This chapter introduces the bases of the next chapter, where the properties of the sound source localization are analyzed.

In order to achieve the goal of the thesis -the necessity of generating acoustic maps for the detected object representation and once the foundations of the auditory system and auditory factors which contribute to the sound source localization have been presented, two methods for the human spatial hearing and sound localization in situations involving multiple sound sources are explained in the Chapter four. The developed method is based on the application of the non-individual Head-Related Transfer Function to static and moving sound source localization through headphones. Against other methods based on the sound source localization using the non-individual Head-Related Transfer Functions, the approach developed in this thesis is based on the study of the evolution of the time delay between two characteristic sounds and its importance in the sound source localization through headphones. Unlike other methods, which analyse the sound source localization and sound parameters directly in the anechoic chamber, where the users localize the sounds delivered by the system through speakerphones or headphones, the proposed approach analyses the sound source localization and its parameters in off-line. The Head-Related Transfer Functions are calculated and measured using a KEMAR manikin and later convolved with a sound through a computer program.

Two sets of experiments are described according to the examined spatial performance involving simple broad-band stimuli. Both experiments measured how well single and train of static and moving sounds are localized in laboratory conditions, for future implementation in the navigation system. These experiments demonstrated that sound source is essential for accurate three-dimensional localization. The approach was based on presenting the sounds as overlapped in time, in order to observe the performance in localization; the objective was to see how time delay between two sounds (inter-click interval) influences on sound source localization. It was found that better localization performance was achieved for trains of sounds. Moreover, the sound perception threshold was studied. In the second study the localization of a moving sound source both in distance and azimuth was analyzed. The results demonstrate that the best results were achieved for an inter-click interval of 150ms. When comparing the localization accuracy in distance and azimuth, better results were obtained in azimuth. Also, it was noted that spatial cues such as interaural time difference and interaural level difference play an important role in spatial localization. The interaural cues arise due to the separation of the two ears, and provide information about the lateral position of the sound.

In the fifth chapter a series of experiments was conducted on blind people, in order to measure their performance in object detection and localization involving one or multiple sound sources. The general approach was to present various objects in order to observe the performance in object detection and sound externalization in various situations. Furthermore, the navigation and object detection and localization was tested in different scenarios. The object localization through acoustical signals arises due to the understanding of listened sounds externalization, which provides information about the spatial position of the source.

Three sets of experiments and two preliminary tests are developed; they make use of the acoustical object detection and navigation system. The two preliminary tests carried out the performance of the sound and object localization accuracy, with the mean to see how the end-user manage the system, how he perceive the system functionality, to analyze which components of the system must be improved or changed for a better functionality and usage. After performing the preliminary test the acoustical module was improved and three sets of experiments were developed: In the first one, a group of seven exercises with different levels of complexity were carried out. It was found that the blind users were able to externalize the sounds provided by the system with a great accuracy and localize the environmental objects. In the second experiment, the navigation task was analyzed. A scenario based on eight soft objects placed at 2,5m of distance creating a labyrinth was used. During the experiment, remarkable results on object detection and localization were observed, despite simultaneous sounds. Slight errors on navigation accuracy were observed when subjects were navigating through the trajectory. The detection task was successfully carried out; with regards to the localization of the objects, the users perceived small deviations on object lateral localization, i.e., some users had difficulties in detecting the object volume. When forcing the user to pass between the objects, the objective was to localize the objects and to avoid them. The errors appearing in that case could be explained by interference of the reproduction of multiple sounds representing different objects, having the users to precise where the location of each one of the object was. On the third set of experiments the individuals were navigating through controlled and uncontrolled outdoor scenarios (playground of a blind people school, street with crossings, bars, restaurants, parking, kiosks, etc). Despite external noises produced by the environmental objects such as music, people speaking, cars noises, etc…, great results were obtained on object detection and localization, and navigation accuracy. The users were able to avoid all objects and navigate with confidence through such complex environments.

In general terms, all these experiments demonstrated that the acoustical representation of the environment is one of the best methods for navigation. It was proven once again that the blind people have considerably great abilities for perceiving the surrounding environment through the hearing. They are able to quickly adapt to the system and to use it as a complementary navigation tool. The acoustical navigation system gives them more confidence and security in navigation. It gives more information from the surrounding environment, information that the white cane is not able to detect. Due to the selected sounds, the system did not interfere with the external noises. Despite all the advantages of the acoustical navigation system, the navigation accuracy depends on the trainings and practice with the use of the device rather than on the sound. All results are based on the end-user feedback giving us directions for refinements, changes, and future work.

The work developed within the context of this thesis has led to the following publications:

Journal papers:

1. Dunai L., Peris F. G., Garcia B.D., Santiago P. V., Dunai I. (2010) “The influence of the inter-click interval on moving sound source

localization for navigation systems". Applied Physics Journal, 56 (3), pp. 370-375
2. Dunai L., Peris F. G., Defez B. G., Ortigosa A.N., Brusola S F. (2009). "Perception of the sound source position", Applied Physics Journal, 55 (3), pp. 448-451

International conference papers:

1. Peris F. G., Dunai L., Santiago P. V., Dunai I. (2010). "CASBliP - a new cognitive object detection and orientation aid system for blind people", CogSys2010 Conference, Zurich
2. Nuria Ortigosa, Samuel Morillas, Guillermo Peris-Fajarnés and Larisa Dunai. (2010), Disparity maps for free path detection, VISAPP 2010 Conference
3. Dunai L., Peris F G., Defez B. G., Ortigosa A.N., (2009). "Acoustical Navigation System for Visual Impaired People", LivingAll European Conference
4. Ortigosa A. N., Dunai L., Peris F. G., Dunai I., Santiago P. V. (2009). "A multiple sensory prototype for visually impaired subject mobility assistance using sound map generation", LivingAll European Conference
5. Santiago P. V., Ortigosa A.N., Dunai L., Peris F. G., (2009). "Cognitive aid system for blind people (CASbliP)", INGEGRAF 2009 Conference
6. Ortigosa A. N., Dunai L., Peris. F.G., (2008). Sound map generation for a prototype blind mobility system using multiple sensors". ABLETECH 08 Conference
7. Fernandes T. M.M., Peris F.G., Dunai L., Redondo J. (2007). "Convolution application in environment sonification for blind people" VII Applied mathematics workshop Valencia
8. Dunai L., Peris F.G., Fernandes T.M.M., Oliver M.J. (2007). "Spatial sound localization base don Fourier Transform", VII Applied mathematics workshop Valencia
9. Javier Oliver, Alberto Albiol, Guillermo Peris, Larisa Dunai. (2007). "HOG descriptor improvement in person detection by means of the reduction of the space dimensions"., Proceedings of VIII Jornadas de Matemáticas Aplicada, UPV, Spain

Key words: blind people, independent navigation, object detection, object motion, sound source perception, inter-click interval.

Resumen

Las personas invidentes y con discapacidad visual han demandado durante muchos años un dispositivo que haga posible una cierta independencia en su movilidad. El diseño y desarrollo de un dispositivo de navegación como el citado supondría un gran hito en el campo de la ingeniería y de la investigación en general. En este contexto, durante estas últimas décadas, diversos investigadores han profundizado en diferentes métodos de representación del entorno de cara a su implementación en dispositivos electrónicos que faciliten la movilidad a personas invidentes y con problemas visuales.

La presente tesis propone el diseño, modelación, implementación, experimentación y análisis de un dispositivo de navegación y detección de obstáculos fácil de utilizar, ideado para personas invidentes. Este dispositivo lleva el nombre de Sistema de Asistencia Cognitivo para las Personas Ciegas (Cognitive Aid System for Blind People – CASBliP en inglés). El dispositivo CASBliP constituye un sistema Electrónico de Ayuda a la Movilidad (Electronic Travel Aid – ETA en inglés), cuyo objetivo principal es ayudar a las personas invidentes a moverse independientemente y de forma segura en diferentes entornos, tanto interiores como exteriores.

En este contexto, la tesis se inicia con la elaboración de un detallado estado del arte sobre los diversos dispositivos de navegación existentes y en desarrollo, destinados a personas invidentes. La revisión efectuada abarca dispositivos desarrollados desde la Segunda Guerra Mundial, momento en el que la construcción de este tipo de dispositivos empezó a jugar un papel más importante en la vida diaria, hasta hoy en día. En este capítulo inicial, se realiza una clasificación de los sistemas de navegación en base al tipo de dispositivo. Más de cuarenta equipos diferentes son descritos en este capítulo. No obstante, a pesar del conocimiento y utilización creciente de los dispositivos de navegación basados en sensores, todavía no ha sido posible el desarrollo de un sistema universal de navegación y detección de objetos, que posea una precisión suficientemente elevada.

Con el fin de lograr los objetivos planteados en la tesis, el dispositivo ha sido diseñado de forma que incluye dos sistemas de entrada, una salida, un portátil y un FPGA como unidades de procesamiento, que pueden funcionar independientemente. El sistema de entrada está basado en una matriz de sensores CMOS Time of Flight de 64×1, implementados en unas gafas y dos cámaras estéreo montadas en un casco. El sistema de salida esta compuesto por un par de auriculares estéreo, a través de los cuales el usuario percibirá los objetos y pasillos libres del entorno. El objetivo del sistema de entrada es

capturar la información del entorno en la dirección frontal al usuario. A partir de la información 3D del entorno percibida por el sistema de entrada, se crean los objetos en movimiento, se detectan todos los objetos móviles y estáticos y los pasillos libres, utilizando para ello los mapas de profundidad, los algoritmos de segmentación y los algoritmos de detección de movimiento. La imagen de alta resolución del sistema de entrada es proyectada en sonidos acústicos de alta calidad, a través del método basado en los criterios de navegación y modelos de convolución con la denominada Función de Transferencia Relativa a la Cabeza (Head-Related Transfer Function HRTF, en inglés).

El dispositivo implementa un método de simulación que es capaz de generar una serie de sonidos a partir de objetos del entorno, de forma que estos sonidos sean capaces de representar la información del entorno con elevada precisión. La idea es generar en el usuario una percepción correcta de las fuentes sonoras virtuales emitidas por la superficie de los objetos, de forma que el cerebro humano se pueda crear una imagen perceptual en tres dimensiones de los objetos, como éstos aparecen en el mundo real. Utilizando esta idea, se pretende crear una percepción global del sonido, permitir a las personas invidentes percibir y crearse una imagen global del entorno circundante, así como el mapa de cómo están organizados los objetos en tiempo real.

Es bien sabido que los seres humanos utilizan una gran variedad de información para la navegación en el entorno, que obtienen a través de la vista, el olfato, el oído, etc. Cuando se daña el sistema de visión humano (ceguera o pérdida parcial de visión), el sistema auditivo toma el mando en lo que respecta a la navegación. En este caso, es muy importante y necesario analizar y definir los aspectos que definen la escena visual, ya que constituyen las características más importantes para la navegación y la detección de objetos, con el fin de representar la presencia de los éstos y determinar su posición en el espacio.

En el capítulo tres se describe, en líneas generales, el sistema auditivo, haciendo referencia a sus componentes básicas, así como a la organización auditiva. Esto proporciona una información preliminar sobre localización de sonidos mediante los parámetros acústicos (parámetros monaurales y biaurales, diferencia interaural de tiempo y diferencia interaural de nivel, efecto de reverberación, cono de confusión, efecto de precedencia y modelo de correlación cruzada). Esta introducción pretende dar una idea general del nivel de precisión necesario en la localización de sonidos, así como introducir la base del capítulo siguiente, en el que se analizan las propiedades de la localización de sonidos.

Para la consecución de los objetivos de la tesis, es preceptiva la creación de un mapa acústico para representar los objetos detectados; en esta línea, a partir de los fundamentos básicos de funcionamiento del sistema auditivo y del estudio de los factores auditivos que contribuyen en la localización de fuentes sonoras, en el capitulo cuatro se describen dos métodos para la audición espacial humana y localización de sonidos en el caso de múltiples fuentes sonoras. El método aplicado se basa en la aplicación de HRTFs no individuales para localización de sonidos estáticos y en movimiento a través de auriculares. Frente a otros métodos existentes, basados en la localización de fuentes sonoras mediante HRTFs no individuales, el enfoque empleado en la tesis se basa en el estudio de la evolución de la característica del tiempo entre dos sonidos y su importancia en la localización de fuentes sonoras a través de auriculares. La función HRTF se calcula y se mide utilizando un maniquí KEMAR y después se convolucionan con los sonidos a través de un software, siendo finalmente ensayados con sujetos reales.

En referencia a las propiedades de localización de sonidos espaciales se describen dos conjuntos de experimentos con sonidos simples de banda ancha. Los dos experimentos analizan la precisión de localización de un sonido y de un tren de sonidos en condiciones de laboratorio, de cara a una posterior implementación en el sistema de navegación. Estos experimentos demuestran que la fuente de sonido resulta clave para la localización tridimensional. La idea consiste en presentar los sonidos desplazados en el tiempo y ver cómo el intervalo temporal entre dos sonidos influye en su localización. Se ha probado que con los trenes de sonidos se obtienen mejores resultados en localización de fuentes sonoras que para el caso de un sonido simple. Asimismo, se analiza también el límite de percepción. En el segundo estudio, se analiza la localización de un sonido en movimiento, tanto en distancia como en azimut. Los resultados obtenidos demuestran que para un intervalo de tiempo de 150ms, se consigue una mejor localización de sonidos. Conviene resaltar que si se comparan la precisión en distancia y azimut, se obtienen mejores resultados en azimut. También se ve en este capítulo que las diferencias interaurales en tiempo y nivel juegan un papel muy importante en la localización espacial. Los parámetros interaurales aparecen debido a la separación de los oídos humanos, que proporciona información sobre la posición lateral del sonido.

En el capitulo cinco se desarrollan una serie de experimentos con personas invidentes, con el propósito de medir su eficiencia en cuanto a la detección de objetos y su localización, cuando intervienen distintas fuentes sonoras. El objetivo general de tales ensayos es presentar diferentes objetos con el fin de observar la habilidad del usuario para la detección de objetos y para la externalización de sonidos en distintas situaciones. La localización de objetos vía señales acústicas se consigue debido a la asimilación del proceso de externalización de sonidos, que proporciona la información sobre la posición espacial del objeto fuente del sonido.

Se describen tres series de experimentos relativos a la detección de objetos y navegación vía sonidos. En la primera serie de experimentos, se desarrollan un conjunto de siete ejercicios con distintos niveles de dificultad (detección de un obstáculo, detección de dos obstáculos, identificación del hueco entre dos obstáculos, detección de una pared, detección de un obstáculo en frente a una pared, etc…). Se ha probado que los usuarios invidentes son capaces de externalizar con gran precisión los sonidos reproducidos por el sistema de navegación y recibidos vía auriculares, así como localizar los objetos en el entorno real. En el segundo grupo de experimentos, se ha analizado la tarea de navegación. Para lograr este objetivo se ha preparado un escenario consistente en 8 columnas construidas a base de cajas de cartón, separadas una distancia de 2,5m, dispuestas en dos líneas formando un laberinto. De dicho experimento se han logrado notables resultados, tanto en lo referente a la detección de objetos como en navegación, a pesar del elevado número de sonidos reproducidos simultáneamente. Se han detectado pequeños errores en cuanto a la precisión en la navegación mientras los sujetos avanzaban por el trayecto dispuesto. No obstante, la tarea de navegación se ha desarrollado, en términos generales, satisfactoriamente; en lo que respecta a la localización de objetos, los sujetos perciben una ligera desviación en la localización lateral de los mismos, es decir, algunos sujetos han tenido problemas con la determinación del volumen de los objetos. El propósito de forzar al sujeto a circular entre los objetos era comprobar si era capaz de detectar los obstáculos y sortearlos. Los errores pueden tener su explicación en la interferencia causada por la reproducción de múltiples sonidos que representaban los diferentes objetos situados en el área de visión; el sujeto debía detectar cada sonido y precisar de dónde provenía. En el tercer grupo de experimentos, el sujeto tenía que navegar por escenarios controlados y no controlados en un entorno abierto (tales como el patio de una escuela para personas invidentes, intersecciones de calles con bares, terrazas, restaurantes, parkings, kioscos, etc…). A pesar de los sonidos exteriores provenientes de señales de tráfico, coches, música, conversaciones humanas, etc…, se han obtenido grandes resultados tanto en localización de objetos como en la navegación. Los sujetos han sido capaces de detectar y esquivar todos los objetos y navegar con confianza en unos entornos tan complejos como los mencionados.

En general, los experimentos desarrollados han demostrado que la representación del entorno mediante sonidos constituye uno de los métodos de navegación más fiables. Se ha corroborado que las personas invidentes poseen una gran habilidad para percibir el entorno, a través del sistema auditivo. Son capaces de adaptarse rápidamente al sistema y utilizarlo como un sistema de navegación complementario al bastón o al perro lazarillo. El sistema de navegación acústico les proporciona mayor confianza y seguridad en la navegación; el sistema les da mucha más información sobre el entorno, información que el bastón u otros sistemas convencionales no pueden detectar. Debido a la naturaleza de los sonidos seleccionados, el sistema apenas interfiere con sonidos externos. La precisión de la navegación depende del entrenamiento y la práctica con el dispositivo y no de los sonidos. Los resultados obtenidos vienen en buena parte influenciados por el feedback con usuario final, que puede dar ideas muy útiles en cuanto a refinamiento, cambios y posibles mejoras.

Como consecuencia del trabajo desarrollado en la presente tesis, se han obtenido las siguientes publicaciones:

Publicaciones en revista:

1. Dunai L., Peris F. G., Garcia B.D., Santiago P. V., Dunai I. (2010) “The influence of the inter-click interval on moving sound source localization for navigation systems”. Applied Physics Journal, 56 (3), pp. 370-375
2. Dunai L., Peris F. G., Defez B. G., Ortigosa A.N., Brusola S F. (2009). “Perception of the sound source position”, Applied Physics Journal, 55 (3), pp. 448-451

Publicaciones en congresos internacionales:

1. Peris F. G., Dunai L., Santiago P. V., Dunai I. (2010). “CASBliP - a new cognitive object detection and orientation aid system for blind people”, CogSys2010 Conference, Zurich
2. Nuria Ortigosa, Samuel Morillas, Guillermo Peris-Fajarnés and Larisa Dunai. (2010), Disparity maps for free path detection, VISAPP 2010 Conference
3. Dunai L., Peris F G., Defez B. G., Ortigosa A.N., (2009). “Acoustical Navigation System for Visual Impaired People”, LivingAll European Conference
4. Ortigosa A. N., Dunai L., Peris F. G., Dunai I., Santiago P. V. (2009). “A multiple sensory prototype for visually impaired subject mobility assistance using sound map generation”, LivingAll European Conference

5. Santiago P. V., Ortigosa A.N., Dunai L., Peris F. G., (2009). “Cognitive aid system for blind people (CASbliP)”, INGEGRAF 2009 Conference
6. Ortigosa A. N., Dunai L., Peris. F.G., (2008). Sound map generation for a prototype blind mobility system using multiple sensors”. ABLETECH 08 Conference
7. Fernandes T. M.M., Peris F.G., Dunai L., Redondo J. (2007). “Convolution application in environment sonification for blind people” VII Applied mathematics workshop Valencia
8. Dunai L., Peris F.G., Fernandes T.M.M., Oliver M.J. (2007). “Spatial sound localization base don Fourier Transform”, VII Applied mathematics workshop Valencia
9. Javier Oliver, Alberto Albiol, Guillermo Peris, Larisa Dunai. (2007). “HOG descriptor improvement in person detection by means of the reduction of the space dimensions”., Proceedings of VIII Jornadas de Matemáticas Aplicada, UPV, Spain

Palabras clave: personas invidentes, navegación independiente, detección de objetos, movimiento del objeto, percepción de fuente sonora, intervalo entre clics.

Resumen

Les persones invidents i amb discapacitat visual han demandat durant molts anys un dispositiu que faça possible una certa independència en la seua mobilitat. El disseny i desenrotllament d'un dispositiu de navegació com el citat suposaria un gran fita en el camp de l'enginyeria i de la investigació en general. En este context, durant estes últimes dècades, diversos investigadors han aprofundit en diferents mètodes de representació de l'entorn de cara a la seua implementació en dispositius electrònics que faciliten la mobilitat a persones invidents i amb problemes visuals.

La present tesi proposa el disseny, modelació, implementació, experimentació i anàlisi d'un dispositiu de navegació i detecció d'obstacles fàcil d'utilitzar, ideat per a persones invidents. Este dispositiu porta el nom de Sistema d'Assistència Cognitiu per a les Persones Cegues (Cognitive AID System for Blind People – CASBliP en anglés). El dispositiu CASBliP constituïx un sistema Electrònic d'Ajuda a la Mobilitat (Electronic Travel AID – ETA en anglés), l'objectiu principal de la qual és ajudar les persones invidents a moure's independentment i de forma segura en diferents entorns, tant interiors com exteriors.

En este context, la tesi s'inicia amb l'elaboració d'un detallat estat de l'art sobre els diversos dispositius de navegació existents i en desenrotllament, destinats a persones invidents. La revisió efectuada comprén dispositius desenrotllats des de la Segona Guerra Mundial, moment en què la construcció d'este tipus de dispositius va començar a jugar un paper més important en la vida diària, fins hui en dia. En este capítol inicial, es realitza una classificació dels sistemes de navegació basant-se en el tipus de dispositiu. Més de quaranta equips diferents són descrits en este capítol. No obstant això, a pesar del coneixement i utilització creixent dels dispositius de navegació basats en sensors, encara no ha sigut possible el desenrotllament d'un sistema universal de navegació i detecció d'objectes, que posseïsca una precisió prou elevada.

A fi d'aconseguir els objectius plantejats en la tesi, el dispositiu ha sigut dissenyat de manera que inclou dos sistemes d'entrada, una eixida, un portàtil i un FPGA com a unitats de processament, que poden funcionar independentment. El sistema d'entrada està basat en una matriu de sensors CMOS Time of Flight de 64×1, implementats en unes ulleres i dos càmeres estèreo muntades en un casc. El sistema d'eixida esta compost per un parell d'auriculars estèreo, a través dels quals l'usuari percebrà els objectes i camins lliures de l'entorn. L'objectiu del sistema d'entrada és capturar la informació de l'entorn en la direcció frontal a l'usuari. A partir de la informació 3D de l'entorn percebuda pel sistema d'entrada, es construeixen els objectes en moviment, es detecten tots els objectes mòbils i estàtics i els camins lliures,

utilitzant per a això els mapes de profunditat, els algoritmes de segmentació i els algoritmes de detecció de moviment. La imatge d'alta resolució del sistema d'entrada és projectada en sons acústics d'alta qualitat, a través del mètode basat en els criteris de navegació i models de convolució amb la denominada Funció de Transferència Relativa al Cap (Head-Related *Transfer* Function HRTF, en anglés).

El dispositiu implementa un mètode de simulació que és capaç de generar una sèrie de sons a partir d'objectes de l'entorn, de manera que estos sons siguen capaços de representar la informació de l'entorn amb elevada precisió. La idea és generar en l'usuari una percepció correcta de les fonts sonores virtuals emeses per la superfície dels objectes, de manera que el cervell humà es puga crear una imatge perceptual en tres dimensions dels objectes, com estos apareixen en el món real. Utilitzant esta idea, es pretén crear una percepció global del so, permetent a les persones invidents percebre i crear-se una imatge global de l'entorn circumdant, així com el mapa de com estan organitzats els objectes en temps real.

És ben sabut que els sers humans utilitzen una gran varietat d'informació per a la navegació en l'entorn, que obtenen a través de la vista, l'olfacte, l'oïda, etc. Quan es danya el sistema de visió humà (ceguera o pèrdua parcial de visió), el sistema auditiu agafa el comandament pel que fa a la navegació. En este cas, és molt important i necessari analitzar i definir els aspectes que definixen l'escena visual, ja que constituïxen les característiques més importants per a la navegació i la detecció d'objectes, a fi de representar la presència dels estos i determinar la seua posició en l'espai.

En el capítol tres es descriu, en línies generals, el sistema auditiu, fent referència als seus components bàsiques, així com a l'organització auditiva. Açò proporciona una informació preliminar sobre localització de sons per mitjà dels paràmetres acústics (paràmetres monaurals i biaurals, diferència interaural de temps i diferència interaural de nivell, efecte de reverberació, con de confusió, efecte de precedència i model de correlació encreuada). Esta introducció pretén donar una idea general del nivell de precisió necessari en la localització de sons, així com introduir la base del capítol següent, en el que s'analitzen les propietats de la localització de sons.

Per a la consecució dels objectius de la tesi, és preceptiva la creació d'un mapa acústic per a representar els objectes detectats; en esta línia, a partir dels fonaments bàsics de funcionament del sistema auditiu i de l'estudi dels factors auditius que contribuïxen en la localització de fonts sonores, en el capítol quatre es descriuen dos mètodes per a l'audició espacial humana i localització de sons en el cas de múltiples fonts sonores. El mètode aplicat es basa en l'aplicació de HRTFs no individuals per a localització de sons estàtics i en moviment a través d'auriculars. Enfront d'altres mètodes existents, basats en la localització de fonts sonores per mitjà de HRTFs no individuals, l'enfocament utilitzat en la tesi es basa en l'estudi de l'evolució de la característica del temps entre dos sons i la seua importància en la localització de fonts sonores a través d'auriculars. La funció HRTF es calcula i es mesura utilitzant un maniquí KEMAR i després es convolucionen amb els sons a través d'un programari, sent finalment assajats amb subjectes reals.

En referència a les propietats de localització de sons espacials es descriuen dos conjunts d'experiments amb sons simples de banda ampla. Els dos experiments analitzen la precisió de localització d'un so i d'un tren de sons en condicions de laboratori, de cara a una posterior implementació en el sistema de navegació. Estos experiments demostren que la font de so resulta clau per a la localització tridimensional. La idea consistix a presentar els sons desplaçats en el temps i veure com l'interval temporal entre dos sons influïx en la seua localització. S'ha provat que amb els trens de sons s'obtenen millors resultats en localització de fonts sonores que per al cas d'un so simple. Així mateix, s'analitza també el límit de percepció. En el segon estudi, s'analitza la localització d'un so en moviment, tant en distància com en azimut. Els resultats obtinguts demostren que per a un interval de temps de 150ms, s'aconseguix una millor localització de sons. Convé ressaltar que si es comparen la precisió en distància i azimut, s'obtenen millors resultats en azimut. També es veu en este capítol que les diferències interaurals en temps i nivell juguen un paper molt important en la localització espacial. Els paràmetres interaurales apareixen a causa de la separació de les orelles humanes, que proporciona informació sobre la posició lateral del so.

En el capitule cinc es desenrotllen una sèrie d'experiments amb persones invidents, amb el propòsit de mesurar la seua eficiència quant a la detecció d'objectes i la seua localització, quan intervenen distintes fonts sonores. L'objectiu general de tals assajos és presentar diferents objectes a fi d'observar l'habilitat de l'usuari per a la detecció d'objectes i per a l'externalització de sons en distintes situacions. La localització d'objectes via senyals acústics s'aconseguix a causa de l'assimilació del procés d'externalització de sons, que proporciona la informació sobre la posició espacial de l'objecte font del so.

Es descriuen tres sèries d'experiments relatius a la detecció d'objectes i navegació via sons. En la primera sèrie d'experiments, es desenrotllen un conjunt de set exercicis amb distints nivells de dificultat (detecció d'un obstacle, detecció de dos obstacles, identificació del buit entre dos obstacles, detecció d'una paret, detecció d'un obstacle davant a una paret, etc…). S'ha provat que els usuaris invidents són capaços d'externalitzar amb gran precisió els sons reproduïts pel sistema de navegació i rebuts via auriculars, així com localitzar els objectes en l'entorn real. En el segon grup d'experiments, s'ha analitzat la tasca de navegació. Per a aconseguir este objectiu s'ha preparat un escenari consistent en 8 columnes construïdes a base de caixes de cartó, separades una distància de 2,5m, disposades en dos línies formant un laberint. Del dit experiment s'han aconseguit notables resultats, tant pel que fa a la detecció d'objectes com en navegació, a pesar de l'elevat nombre de sons reproduïts simultàniament. S'han detectat xicotets errors quant a la precisió en la navegació mentres els subjectes avançaven pel trajecte disposat. No obstant això, la tasca de navegació s'ha desenrotllat, en termes generals, satisfactòriament; pel que fa a la localització d'objectes, els subjectes perceben una lleugera desviació en la localització lateral dels mateixos, és a dir, alguns subjectes han tingut problemes amb la determinació del volum dels objectes. El propòsit de forçar al subjecte a circular entre els objectes era comprovar si era capaç de detectar els obstacles i sortejar-los. Els errors poden tindre la seua explicació en la interferència causada per la reproducció de múltiples sons que representaven els diferents objectes situats en l'àrea de visió; el subjecte havia de detectar cada so i precisar d'on provenia. En el tercer grup d'experiments, el subjecte havia de navegar per escenaris controlats i no controlats en un entorn obert (com ara el pati d'una escola per a persones invidents, interseccions de carrers amb bars, terrasses, restaurants, pàrquings, quioscos, etc…). A pesar dels sons exteriors provinents de senyals de ciruculació, cotxes, música, conversacions humanes, etc…, s'han obtingut grans resultats tant en localització d'objectes com en la navegació. Els subjectes han sigut capaços de detectar i esquivar tots els objectes i navegar amb confiança en uns entorns tan complexos com els mencionats.

En general, els experiments desenrotllats han demostrat que la representació de l'entorn per mitjà de sons constituïx un dels mètodes de navegació més fiables. S'ha corroborat que les persones invidents posseïxen una gran habilitat per a percebre l'entorn, a través del sistema auditiu. Són capaços d'adaptar-se ràpidament al sistema i utilitzar-lo com un sistema de navegació complementari al bastó o al gos guia. El sistema de navegació acústic els proporciona major confiança i seguretat en la navegació; el sistema els dóna molta més informació sobre l'entorn, informació que el bastó o altres sistemes convencionals no poden detectar. A causa de la naturalesa dels sons seleccionats, el sistema quasi no interferix amb sons externs. La precisió de la navegació depén de l'entrenament i la pràctica amb el dispositiu i no dels sons. Els resultats obtinguts vénen en bona part influenciats pel *feedback* amb usuari final, que pot donar idees molt útils quant a refinament, canvis i possibles millores.

Com a conseqüència del treball desenrotllat en la present tesi, s'han obtingut les publicacions següents:

Publicacions en revista:

1. Dunai L., Peris F. G., Garcia B.D., Santiago P. V., Dunai I. (2010) "The influence of the inter-click interval on moving sound source

localization for navigation systems". Applied Physics Journal, 56 (3), pp. 370-375

2. Dunai L., Peris F. G., Defez B. G., Ortigosa A.N., Brusola S F. (2009). "Perception of the sound source position", Applied Physics Journal, 55 (3), pp. 448-451

Publicacions en congressos internacionals:

1. Peris F. G., Dunai L., Santiago P. V., Dunai I. (2010). "CASBliP - a new cognitive object detection and orientation aid system for blind people", CogSys2010 Conference, Zurich
2. Nuria Ortigosa, Samuel Morillas, Guillermo Peris-Fajarnés and Larisa Dunai. (2010), Disparity maps for free path detection, VISAPP 2010 Conference
3. Dunai L., Peris F G., Defez B. G., Ortigosa A.N., (2009). "Acoustical Navigation System for Visual Impaired People", LivingAll European Conference
4. Ortigosa A. N., Dunai L., Peris F. G., Dunai I., Santiago P. V. (2009). "A multiple sensory prototype for visually impaired subject mobility assistance using sound map generation", LivingAll European Conference
5. Santiago P. V., Ortigosa A.N., Dunai L., Peris F. G., (2009). "Cognitive aid system for blind people (CASbliP)", INGEGRAF 2009 Conference
6. Ortigosa A. N., Dunai L., Peris. F.G., (2008). Sound map generation for a prototype blind mobility system using multiple sensors". ABLETECH 08 Conference
7. Fernandes T. M.M., Peris F.G., Dunai L., Redondo J. (2007). "Convolution application in environment sonification for blind people" VII Applied mathematics workshop Valencia
8. Dunai L., Peris F.G., Fernandes T.M.M., Oliver M.J. (2007). "Spatial sound localization base don Fourier Transform", VII Applied mathematics workshop Valencia
9. Javier Oliver, Alberto Albiol, Guillermo Peris, Larisa Dunai. (2007). "HOG descriptor improvement in person detection by means of the reduction of the space dimensions"., Proceedings of VIII Jornadas de Matemáticas Aplicada, UPV, Spain

Paraules clau: persones invidents, navegació independent, detecció d'objectes, moviment de l'objecte, percepció de font sonora, interval entre clics.

Thesis objectives

Nowadays none of the developed Electronic Travel Aids is capable to make the blind users and visually impaired people to feel more confident. This is mainly due to the limitations of the currently available technologies and methodologies for environmental information acquisition processing and reproduction.

The present thesis objective is to develop a complete working mobility device for blind people, which will capture environmental information via sensors and stereo-cameras transforming it into acoustic signals. To define the basis of the wearable obstacle detector and navigation device, understanding by wearable, the possibility of combining in a suitable way a non-prohibitive cost for the blind users, the device functionality and its accuracy.

In order to achieve the main aim of the thesis, several specific objectives were carried out:

- To develop a robust device, by using the cutting-edge technology
- To research on the blind people needs and abilities for navigation and environmental information perception
- To determine the types of environmental information needed for acquisition and display and to study all possible combinations of environmental scenes which might appear in the navigation space of the user
- To design simple technology for environmental information acquisition and processing. To develop the sensor and camera stabilization and object detection methodology, the motion detection methodology (it is important to take into account that the moving objects located in the near environment of the user are crucial for him), the acoustic representation of the processed information received by the cameras and/or sensor.

  For acoustical representation of the environment, it is necessary to discuss about the development of the acoustic interface and the use of the localization methodology. The objective of the acoustic task requires:
- To investigate the acoustic sounds, to understand the auditory spatial perception and the human auditory system
- To select the type of sounds which will be used for the final system; it is important to select three types of sounds, because the system interprets by three different sounds the objects detected by the sensor, the objects detected by the stereo-camera and the free paths.

All three sounds must neither mutually interfere nor interfere with the external sounds.

- To research the acoustic parameters of the spatial sound and the localization task via headphones in laboratory conditions with pure tones and noisy environment

And, finally, it is an important to test the navigation and object detection task with blind users by collecting the feedbacks and by analysing the statistical results, using them as a starting point for future improvements.

Research contribution and innovative aspects of the system

The main purpose of the Cognitive Aid System for Blind People is to assist blind users towards their independent mobility both in indoor and in outdoor environments, complementing their main mobility aids, i.e., the traditional white cane, guide dog or GPS. The most important environmental elements that affect the mobility/navigation of the blind and sighted people are the obstacles and the free paths. Usually the farther environmental obstacles cannot be detected by the white cane, thus the ETA systems come in help to the blind user. The device is equipped with an obstacle and free path detection functionality which provides a more complete perception of the 3D environment (real environment) including elements of special interest for the blind user, such as moving and static objects and free paths. The CMOS Time of Flight and Stereo Camera system constitutes the main unit of the system. Using modern algorithms, we obtained image depth maps (3D structure of the environment) and object motion (moving objects). The second important feature of the system is the acoustic system consisting of a collection of spatial sounds covering an area up to 15m in distance and 64º in azimuth.

Among the novel methodologies which have been used in the thesis, we can remark: the object and/or free path detection, motion detection methods, image resolution processing and output image resolution (acoustic resolution) and, finally, the acoustical map developed with three different acoustical signals convolved with non-individual Head related Transfer Functions.

CMOS Time of Flight sensor developed by SIEMENS is a laser used for traffic. The innovative aspect is its minimisation and its use on navigation task for blind people. A major characteristic of the CMOS Time of Flight sensor is its detection accuracy, (the laser has and accuracy of 99%) and processing time (real time). An array of 64×1 sensors have been implemented into a pair of glasses in the horizontal plane.

Moving object detection – a very important issue for the blind people, with regards to safety navigation, lies in the moving objects (people, cars, bikes, etc.); these objects must be early detected (before they may penetrate in the cane detection range). This task was achieved by implementing the inertial sensor on the hardware and software for motion algorithm.

Free path detection – a detection method using depth discontinuities. The proposed method exploits this by looking for linear variation in the depth map.

Image resolution – the output of the stereo vision method is based on high resolution images (1024x768) and the auditory image is 15m ×64º of externalized sounds. In our case the motion is presented by means of the pitch modification method. The developed method takes into consideration navigation criteria: distances of the obstacles, free paths, static and moving objects and objects of interest (the most dangerous objects including the nearest and fastest objects). This means that when the object gets nearer with respect to the user, the inter-click interval can decrease from 100ms up to 25ms.

Acoustic output/interface – three very short acoustical sounds are used for representing the environment: a delta sound is used for object detection through the sensor system, a Synthetic Percussive sound for the moving objects and a Modified Synthetic Musical sound for the free path coordinates. These sounds were convolved with non-individual Head Related Transfer Functions (HRTFs) in order to obtain the spatial mean of the sounds. The spatial position of the object is given by the coordinates x (azimuth) and y (elevation) corresponding to the centre of their bounding boxes. The distance, or z coordinate, is represented by the closer coordinate provided by the system. In addition, the pitch of the sound changes progressively with the distance. Trains of sounds generated and processed with different timbres, pitches, and inter-click intervals (ICI) are used for different objects representation.

Experimental methodology – an additional contribution of the thesis consists of the integration of all above mentioned methodologies into a single integrated method. In particular, the object detection and object representation constrained by the technological limitations. Beside the experimental methodology established for the navigation task, several methods for the psychoacoustic experiments were developed in order to research the sound source localization. In this task, the sound source localization was studied for far fields. Also, the influence of the ICI and its threshold on the sound source localization was analyzed.

The overall advantage of the research is the **final product**, to get a unique prototype methodology for further studies, feedback from users testing different scenarios, feedback from the users for improving the image resolution and acoustic module, creating new knowledge of the 3D sensation of the environment for the blind people.

Declaration

As a consequence of this work, the following publications have arisen:

Publications in indexed journals JCR

1. Dunai L., Peris F. G., Garcia B.D., Santiago P. V., Dunai I. (2010) “The influence of the inter-click interval on moving sound source localization for navigation systems”. Applied Physics Journal, 56 (3), pp. 370-375
2. Dunai L., Peris F. G., Defez B. G., Ortigosa A.N., Brusola S F. (2009). “Perception of the sound source position”, Applied Physics Journal, 55 (3), pp. 448-451

Publications in international and national conferences

1. Peris F. G., Dunai L., Santiago P. V., Dunai I. (2010). “CASBliP - a new cognitive object detection and orientation aid system for blind people”, CogSys2010 Conference, Zurich
2. Nuria Ortigosa, Samuel Morillas, Guillermo Peris-Fajarnés and Larisa Dunai. (2010), Disparity maps for free path detection, VISAPP 2010 Conference
3. Dunai L., Peris F G., Defez B. G., Ortigosa A.N., (2009). “Acoustical Navigation System for Visual Impaired People”, LivingAll European Conference
4. Ortigosa A. N., Dunai L., Peris F. G., Dunai I., Santiago P. V. (2009). “A multiple sensory prototype for visually impaired subject mobility assistance using sound map generation”, LivingAll European Conference
5. Santiago P. V., Ortigosa A.N., Dunai L., Peris F. G., (2009). “Cognitive aid system for blind people (CASbliP)”, INGEGRAF 2009 Conference
6. Ortigosa A. N., Dunai L., Peris. F.G., (2008). Sound map generation for a prototype blind mobility system using multiple sensors”. ABLETECH 08 Conference
7. Fernandes T. M.M., Peris F.G., Dunai L., Redondo J. (2007). “Convolution application in environment sonification for blind people” VII Applied mathematics workshop Valencia

8. Dunai L., Peris F.G., Fernandes T.M.M., Oliver M.J. (2007). “Spatial sound localization base don Fourier Transform”, VII Applied mathematics workshop Valencia
9. Javier Oliver, Alberto Albiol, Guillermo Peris, Larisa Dunai. (2007). “HOG descriptor improvement in person detection by means of the reduction of the space dimensions”., Proceedings of VIII Jornadas de Matemáticas Aplicada, UPV, Spain

Table of contents

Tables

Figures

Abbreviations

ACER - laptop model
AWT – Absolute Walking Time
BLP – Basic Learning Protocol
CASBliP – Cognitive Aid System for Blind People
CCD – fireware colour camera
CIDAT – Centro de Investigación, Desarrollo y Aplicación Tiflotécnica
CMOS – complementary metal-oxide-conductor
C-4 – Laser came model C4
C-5 – Laser Cane model C5
C++ - programming language
DBSV – Deutscher Blinden- und Sehbehindertenverband e.V.
DLL – dynamic-link library
DSP - Digital signal processing
ETA – Electronic Travel Aid
EPFL - Swiss Federal Institute of Technology
ESA – European Space Agency
EXIT – exit
FPGA – Field Programmable Gate Array
FSRS – Forehead Sensory Recognition System
FL2-03S2 – Time of Flight sensor model
GaAs – Galium Arsenide
GIS – Geographic Information system
GPS – Global Positioning System
HRTF – Head-Related Transfer Function
HD-201 – headphones model
HPS – Head Position System
ICI – inter-click interval
IEEE1394b – serial bus interface standard
IFC – Istituto dei Ciechi “Francesco Cavazza”
ILD – Interaural Level Differences
ITD – Interaural Time Differences
KEMAR – Kemar manikin
kHz – Kilo Hertz (measure unit)
L – Left
LAB_EXP – laboratory experiment
M – Meter
Min – minutes
MLBS – Maximum Length Binary Sequence
MMC – multimedia card reader

MoBIC – Mobility of Blind and Elderly People Interacting with Computers
MoPs – MoBIC Pre-Journey System
MoODS – MoBIC Outdoor System
M1 – Acoustic Prototype
M2 – Real-Time acoustic Prototype
NAVI – Navigation Assistance for Visually Impaired
NC – number of correction
NH – number of heats
NRC – National research Council
NOD – Nothingham Obstacle Detector
PDA – Personal Digital Assistance
PGS – Personal Guidance System
PC - computer
P1-P9 – subject identification
PA, PB – point A and point B
RIAS – Remote Infrared Audible Signage
RFID – Radio Frequency Identification
R – Right
SBPS – single Board Processing System
SONA – sonic Orientation Navigation Aid System
SWAN – System for Wereable Audio Navigation
START – initialization operation
STOP – finalization operation
SONY MDR-EX75SL – headphone model
TP – subject identification
TANYA – Tactile Acoustical Navigation and Information assistance
VAS – Virtual Acoustic Space
vOICe – Seeing with sounds project
USB – Universal Serial Bus
º - degree
2D – two dimensional
3D – three dimensional

1. CHAPTER I: NAVIGATION SYSTEMS

CHAPTER I: NAVIGATION SYSTEMS

1.0.Summary

In this chapter, a detailed state-of-the art of the different commercial and prototyped navigation systems developed from the 50’s is carried out. Also the motivations for the development of the Electronic Travel Aids and the user requirements are analyzed. The objective is to show the important advances developed in this field, which has drawn increasing concern during this last decades. Special attention is paid to their adaptation to people with visual deficiencies, since this is an area in which the potential advantages of these devices become particularly relevant. We show that significant advances have been made in the development and dissemination of electronic travel aids using objective and subjective, direct and indirect measurements in the mobility training programs for blind people.

Regarding the blind user requirements, there is an urgent need to determine how and which information has priority to be acquired from the environment for the navigation task. Also, the importance of the visual sensory substitution by the enhanced technique which can be used to compensate the visual deficiency and the information representation methods is described. Due to the fact that the human perceiving abilities decrease with the distance, the design and development of output artificial devices play an important role on electronic travel aids. That is why for the output devices various methods, such as auditory, tactile or vibration perception have been developed.

A classification of the mobility aid systems into three main groups is established: ‘obstacle detectors’, ‘environmental sensors’ and ‘navigation systems’. The main differences between those categories as well as the most relevant devices included in each one are described and commented.

The characteristics of the different devices described in the chapter have been obtained from the available publications in different journals and conferences, which are included in the references section. Moreover, a special section including the most relevant projects funded by international organisms and whose aim is the development of this type of devices has been included in the chapter.

1.1 Motivation

Information in the environment enables humans and vertebrates to learn about sources that are in many different directions, particularly signals that are outside the detection range of other senses (Fay and Popper, 2005). Sound source localization is inherently important for safety-survival and navigation. Blind people make maximum use of sound not only to know the obstacle presence but also where is and how dangerous is in order to avoid it effectively.

There are over 314 million of blind and partially sighted people in the world from where 45 millions are total blind (WBU, 2009). Blindness is the condition of lacking visual perception due to physiological or neurological factors.

There are several main skills that the blind community requires:

1. live independently and productively
2. communication
3. raise a family
4. have a social life
5. mobility
6. maintain a career- or launch a new one
7. enjoy sports, games

Loss of vision often is accompanied by loss of independence. Visual impaired and, in particular, total blind people are unable to take advantages of different services. They have lack of social interaction, human contact and they are limited in mobility.

Communication is an important factor in the human life. Reading, writing and speaking make humans more communicative, friendly and help them to interchange knowledge's, explain situations and feelings, happiness and sadness. The communication helps people to enrich intellectually and spiritually.

Nowadays various techniques are developed for reading and writing for blind community: Braille, talking books, reading machine which convert the printed text into speech or Braille. Also a variety of computer software and hardware such as mobiles, scanners and refreshable Braille display, optical character recognition applications and screen readers, radio reading services etc., help blind community to communicate with the surrounding people, familiars and unfamiliar people.

One of the main necessities of the blind people is the lack of mobility, which become a severe constraint for the person. Blind people find difficulties to travel independently, because they can not determine their positions and objects location in the surrounding environment. For the sole purpose of getting out a considerable amount of information is required.

Loose of communication and mobility forbid blind people to enjoy sports, tracking, to obtain a career or a good job. These limitations make blind people be dependent all their life. They can not take the public transport or make shopping. They always need somebody to accompany them.

Many ways and technologies have been introduced to help blind community, in order to satisfy the basic desire of mobility and communication.

Blind Unions help blind users to learn to use various techniques and methods of reading, writing and navigate. Also learn how to improve other body part which will help them to orientate and perceive the surrounding. They learn to make use of the sounds, feelings, temperature, etc. to help them in their habitual life.

Most of blind and partially sighed people learn to use their audition to compensate the lack of vision. Environmental information enables the humans and animals to learn about sources and sounds from the surrounding.

Acoustic information is a primary tool for orientation by blind and partially sighted people, for example, to determine when traffic has actually stopped – rather than when it has been signaled to stop. And when crossing at an intersection that has no traffic lights, they listen for oncoming traffic to determine when to cross.

Blind people make maximum use of the sounds in order to maintain its safety.

- For designing and building more effective mobility devices, it is important to carry out an analysis of the size, needs and characteristics of the population to be served.

- The assessment of mobility is necessary for detecting the success and failure of the mobility tests and effectiveness of the electronic travel aids. Due to multiple characteristics, such as psychosocial, sensory and motor, perceptual and cognitive, and environmental characteristics, the individual mobility performance is affected.

- Since a great part of the sever visually impaired people suffer of at least one additional impairment, it is advisable to develop a study focused on the interaction of visual impairments with other cognitive, sensory, and physical impairments and their effect in the mobility problems. Table 1.1 represents the main characteristics that affect the mobility.

- New devices development often leads to difficulties of usage and working principle. That is why the development of specific guidelines, research protocols with small samples that could enable a systematic evaluation and recommendations for future scientists becomes important.

- In the older people the prevalence of blindness is greater that in the younger. Due to that, it is advisable that the mobility requirements and preference of elderly visually impaired people to be conducted. It is necessary to bear in mind the variety, severity and distribution of their visual impairment, cognitive or motor diseases that may affect the usage of an Electronic Travel Aid.

Table 1.1 The main characteristics that affect mobility

Psychosocial Characteristic	Individual expectation
	Living situation
	Motivation
	Demographic profile
Sensory/Motor Characteristic	Visual capacity
	Auditory capacity
	Locomotor capacity
Perceptual/Cognitive Characteristic	Ability to usage echoes
	Spatial awareness
	Cognitive capacity/processing mode
Environmental Characteristic	Urban/suburban/rural
	Weather conditions
	Time of day
	Previous training/learning
	familiarity

An important factor for the Electronic Travel Aids evaluation and dissemination is related to measurements methods. Two types of these methods have been hitherto developed; the indirect and direct methods. The indirect measurement method accesses the underlying perceptual and cognitive processes that are responsible for overt behavior. This method is used for enhancing the sensitivity of the measurement providing indices of stress or demands in the mental capacity of the mobility task. The direct measure method measures specific aspects of the mobility, such as obstacle detection, or specialist judgments. There are two methods; the objective and subjective measurement. Direct objective measurements could be found in the Table 1.2. The direct subjective measurement represent the specialists judgments, self-reports of the performances, rating scales applied in the evaluation, personal satisfaction etc.

Table1.2 Direct Objective Measures for mobility

Safety	obstacle contacts
	unsafe veering
	Fail to detect step-down
	Fail to detect step-up
	Inner shoreline body contact
Efficiency	Walking speed
	Continuity of travel
	Navigational errors
	Veer/route departures
Travel Frequency	
Travel distance	
Travel familiarity or Difficulty	
Device/Sensory System Specific Measurement	Sign detection
	Search time
	Reading time
	Surface structure/texture determination
	Drop-off detection
	Directional determination
	Distance determination
Ratio of independent to Sighted Guide Travel	
Object of Landmark Detection	

1.2 Electronic Travel Aid Systems

Foulke (1971) defined the “mobility” as “the ability to travel safety, comfortably, gracefully and independently through the environment”. Basically, there are three ways of getting mobile. The first one is when the blind person uses the sighted human guide. The second and third are the most common and accepted methods, the use of the white cane and the guide dogs. The guide dog is able to detect and analyze complex situations such as crass walks, stairs, potential dangers etc., where the information is received by tactile feedback.

During many centuries the “cane” (see Figure 1.1) has been the most popular mobility aid system. Despite its importance in the blind community, before 1964, when Russell C. Williams published the “Specifiations for the long cane (Typhlocane)”, which helped to establish a long cane model, the used canes lacked of any standards and specifications (Farmer, 1978). The people used any kind of stick and canes to help them into navigation. Through long cane the user perceives the very near environment not farther than 1,5 m. The long cane represents the tactual sense of the user, like an extension of the hand (WBU, 2009). In 1971, during the National Research Council conference in Washington, the physical and functional characteristics of the long cane were adopted (NRC, 1972). These characteristics concerned the crock, grip, tip and the shaft of the long Cane:

1. Vertical axis of shaft must be straight.
2. Slight tapes of shaft from grip to tip.
3. It should be available in various lengths to fit height of individual user.
4. It should have enough length to provide the user with essential information in ample time to react to it but not to inhibit the user’s physical freedom.
5. Its weight should be as light as possible without affecting balance of sacrificing other requirements.
6. It should have a low wind resistance.
7. It should have enough rigidity to enable user to establish accurate distance and position of the detected object; without excessive whip or bend, maintaining original shape under stress.
8. It must not conduct significant amounts of thermal or electrical energy.
9. It should enable an adequate transmission of vibrations from the tip to the grip to provide best tactile and aural stimulus.
10. It should be rigid, visible to pedestrians or conductors.

11. It must produce minimal noise level when used without artificial dampening devices.
12. It should be durable.
13. It should be replaced easily without special tools.
14. It should have good balance.
15. It must have acceptable appearance.

Although the long cane is considered the most effective and efficient mobility device for independent mobility, it has some disadvantages.

Figure 1.1 The long cane

The long canes do not provide enough information when a dangerous obstacle is near or against collisions. The length of the cane limits the range of the environment perception, loosing a substantial amount of information.

After the Second World War, with the sensor development, many efforts have been made to design and develop Electronic devices able to perceive the surrounding environment. In 1897, Starkiewicz and Kuliszewsky built the first Electronic Travel Aid system (ETA), the “Neiszewski‘s Electroftalm”, and in 1912 Nye and Bliss developed the “D’Albe’s Exploring Optophone”. In 1970 Nye and Bliss published one of the best historical reviews on Electronic Travel Aid systems. Farmer (1975) mentioned that an ETA should, itself or with a cane or dog guide, inform the blind traveler on objects in the travel path from the ground to the vicinity of the head, as well as forewarn of any surface discontinuities that might constitute a safety hazard. The Elektroftalm used a single selenium cell placed on the forehead in order to control the sound output intensity (Starkiewicz and Kuliszewski, 1963). With that technology the blind user was able to distinguish between light and dark. The Exploring Optophone developed by D’Albe’s had a very similar operation procedure, converting light into sounds (Capp and Picton, 200), (Meijer, 1992). A selenium detector and a sound delivered to the light failing intensity were used. This device was presented in 1912 at the Optical

Convention in the Science Museum in South Kensington. ETA systems enable blind users to avoid the obstacles in the front of view. They warn about the presence of distant obstacles. They determine the direction, range, width, height and other object particularities. ETA systems constitute a hope for the blind community, enabling blind users to travel with more confidence and security. Travelling with confidence and security helps blind users to move with less tension and stress. Some authors mentioned that ETA systems are designed to be systems complementary to the cane or guide dogs which will help blind users to navigate through environment.

Different theorems about ETA systems have been proposed. Benham (1954) and Benjamin (1968) suggested that an ETA system should detect obstacles, and indicate their approximate location and distance, big or small obstacles, lightweight or hard obstacles, steps or free path. Also, the device, as an additional informational input should give the blind traveler additional orientation and navigation information. This information should not interfere with the environment noises, information which will not give false cues (Dupress, 1963) and be synchronized with other cues. The information provided should be clear and simple. The information should be quickly interpreted without extensive training. For Farmer (1975), an ETA system should be a device which will contribute to the user independence on the mobility and will not hinder him.

Beside the working requirements and system specifications mentioned above, an ETA system should be designed in order to contain the minimum possible number of accessories (boxes, electronics, helmets and connecting cables, etc.,) in order not to bother and disturb the user. Besides the importance of the system accessories, it is also significant the importance of the techniques for representing the information acquired from the environment.

From a technical point of view, the devices should be designed and developed not only to be light and small, but also reliable and durable and esthetically well designed, to have a high quality and to assure reliability during their operation.

Technically, the ETA systems are based on tree interfaces: the input interface, the processing interface and finally the output interface. The input interfaces acquire the environmental data. They can be classified in: ultrasound, laser, artificial vision and GPS systems. The processing interface contains the techniques and software for processing all the acquired information and for transforming it into the required data for the output interface. The output interface is, as the previous interfaces, important. The output system represents the model for transmitting the information from the device to the user. It should be as much concise and clear as possible, in order not to confuse and disturb the user.

Table 1.3 summarizes the main ETA systems as well as the basic specifications. The features of these systems will be commented during the following points.

Table 1.3 Electronic Travel Aids specifications

Nº	ETA name	Specifications
		Ultrasonic
1.	Nottingham Detector	- Obstacle detector device - Hand-held device - Provide pulses of high frequency sound (40 kHz), similar to the major musical scale - 8 outputs - Maximum detection range is 7 feets
2.	Sonic Torch	- Hand-held narrow-beam system - auditory signals output - the pitch corresponds to the range - the timbre corresponds to variations in target surface texture Wide band-width frequency moduled ultrasonic energy wave (40-80 kHz) - angle of view- a cone of 30º on either side of the midline direction of view
3.	Lindsay Russell Pathsounder	- obstacle detector device - range of detection – 6feets ahead the user - tactile and auditory outputs - a box suspended from the user neck
4.	Mowat Sensor	- hand-held system - sonar and tactile output - it emits elliptical ultrasonic cone of 15º wide at 30º high. - Vibrates at a rate inversed to the distance from
5.	Sonic Pathfinder	- head-held device - audio signals outputs
6.	Sonic Guide	- binaural system (is composed by two chanels) - head-mounted device - range of detection – 6m - sound, tactile and vibrotactile output

7.	Walkmate	- object detector system - tonal output
8.	Polaron	- obstacle detector or clear path indicators system - vibrotactile or auditory signals output - detection range – 16 feets
9.	Sensory 6	- object detector device - tonal output
10.	Ultracane (Batcane)	- environmental sensors - object detector system - detect object texture and flow patterns
11.	Light Probes	- photocells connected to the electronic circuit - light source detection - acoustic, speech, tactile, vibratory output
12.	Palmsonar	- photocells and electronic circuits based system - detects the light source - acoustic, speech, tactile and vibration output
	Laser	
13.	Laser cane	- gallium arsenide laser - acoustical signals at the output - detection range of 4 m ahead the user
14.	Pilot Light	- mini-radar - infrared system - hand-held device - detection range 3,5m - talking output
15.	Talking light	- infrared wireless communication - hand-held system - environmental sensors - human voice output
16.	Talking signs	- Infrared wireless communication - RIAS system - Hand-held device - Infrared light - Infrared external sensors
17.	SONA	- environmental system - radio-activated auditory beam
18.	MARCO	- infrared light transmitter and receiver system - hand-held device - detection range 12,5m - human voice output

19.	Verbal Landmarks	- Orientation and navigation system - Inductive loop radio signals - Detection range 2m - Onmidirectional - Human voice output
20.	Easy Walker	- Navigation system - Infrared laser input
21.	CASBliP M1	- object detection device - Time of flight sensor - Distance perception range 5m - acoustical pitch output
	GPS	
22.	MoBIC	- Orientation and navigation system - Hand-held keyboard - Acoustical signals
23.	Makino	- Orientation and navigation device - Digital mobile phone - Computer central unit - Synthetic speech output
24.	Electronic Guide Dog	- Orientation and navigation system - Mobile phone and central unit - Human agent at the central - Hand-held system - Voice communication output
25.	Talking radar	talk
26.	PNG	- Orientation and navigation system - Fluxgate Compas and digital maps, GPS - GIS connection - Synthetic speech and vibrotactile output
27.	GPS Braille Note	- GPS, maps and points of interest - Speech and Braille output
28.	Trekker	- GPS and digital maps - Vocal output
29.	TORMES	- hand-held device - EGNOS technology, GPS and Braille Keyboard - Synthetic speech output - Navigation accuracy 2m

30.	Easy Walk	- GPS, mobile phone, Symbian operating system - Bloetooth GPS receiver - 24 hours call-center - SMS and call to the call-center - Speech output
31.	Geotact	- GPS, inertial sensor - Speach
32.	Talk & Wayfinder	- Navigation device - GPS - Voice-led menu and vocal output
33.	Loadstone GPS	- Mobile phone, Symbian operating system, Bluetooth GPS receiver , USB Bluetooth dongle, Bluetooth headset, Multimedia Card Reader - Voice output - Points of interest
	<u>Artificial vision</u>	
34.	vOICe	- Navigation system - camera - acoustical sounds and tactile maps output
35.	CASBliP M2	- Object detector and navigation system - Two cameras, inertial sensor, PC - static and dynamic object and free path detection - acoustical signal output
36.	EPEL	- Stereoscopic sonar system - Senar sensors - Vibrotactile output
	<u>Combined</u>	
37.	CASBliP	- time of flight sensor, two cameras, inertial sensor, PC - static and dynamic, free path detection - acoustical sound output
38.	SWAN	- Navigation system - tactile, GIS four cameras, GPS, digital compass PC - audio output

39.	TANIA	- hand-held device - maps, sensory system and tactile-acoustical interfaces, PC tablet, MTx inertial sensor, Braille display - numbers, distance recognition - acoustical sounds and display information representation output
40.	Tyflos	- Reading and navigation device - 2 cameras, range sensor, GPS and RFID reader - Vibrations and speech output
41.	Eye PlusPlus (FSRS)	- Videocamera - Vibrotactile output - Glasses and body belt

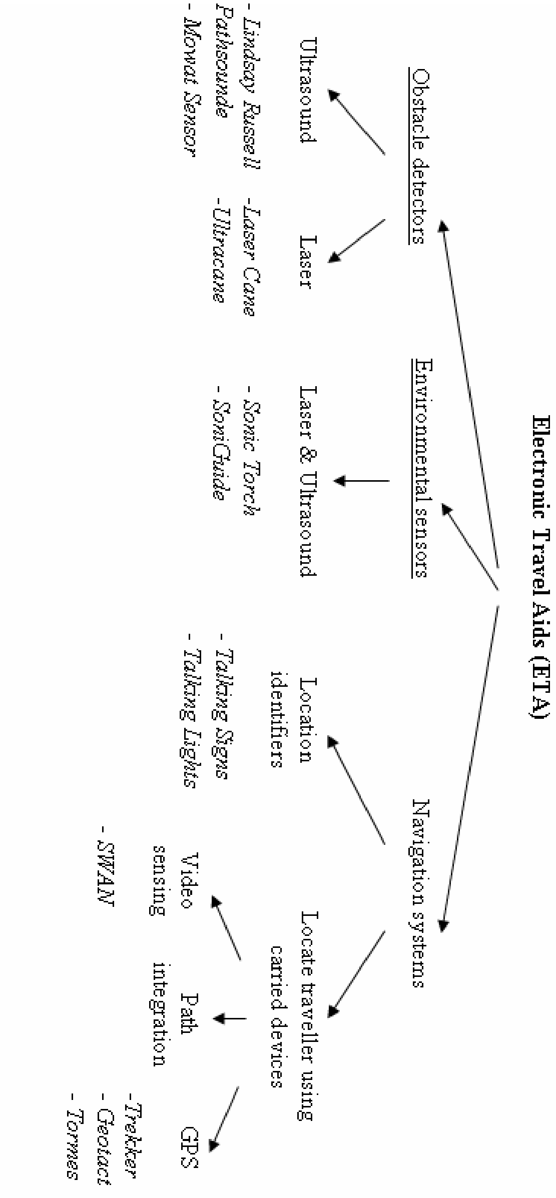


Figure 1.2 Electronic Travel Aids classification

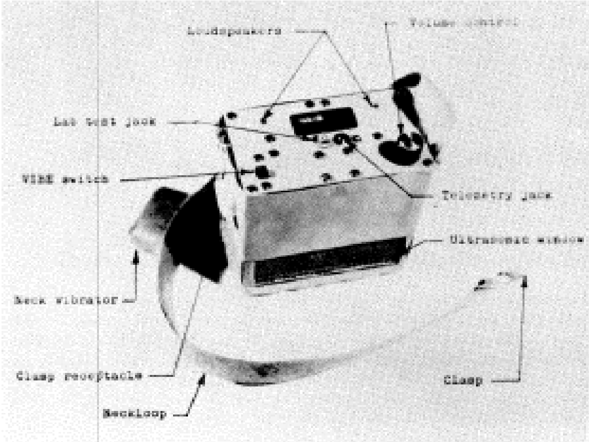


Figure 1.3 Lindsay Russell Pathsounder

Many of the existing electronic travel aids are based on ultrasonic and optical technologies (Brabyn, 1982). Nowadays, more than fifteen different ultrasound devices are known in the whole world. Since early times, the mobility aids have been classified in *tree classes* (see Figure 1.2). The first class is known as *obstacle detectors or clear-path indicators* (Foulke, 1971). These aids warned only about the presence of very extremely near obstacles which were directly in the travel path. These devices are considered as go-no-go systems, which provide limited information. One of these categories is the ***Lindsay Russell Pathsounder*** (Russell, 1965), which is considered the first developed ultrasound mobility device (Mann, 1970). The system uses a 30º ultrasonic beam transmitted from a chest-level unit suspended from the user neck. Russell Pathsounder system (Figure 1.3) can warn about the presence of the objects up to 6 feet ahead with certain range estimation via a tactile and an auditory interface. The system indicates the presence of objects in two similar zones: an outer protection zone and an inner protection zone. A vibratory system was incorporated with intends to serve to the hearing disabled blind person. Also, it is used to replace the auditory signal in a noisy environment (Russell, 1974). The main aim of Lindsay Russell Pathsounder was to complement the currently existing white cane, by providing the blind user protection and giving distant early warnings. Over the years, the Lindsay Russell Pathsounder has been subjected to multiple modifications (Farmer, 1978) for its improvement from the Model E up to the modern Russell Pathsounder.

Various prototypes of the Russell Pathsounder were developed for blind people with different and multiple handicaps. One of these devices is the ***Pathsounder Special E***. This device consists of two units: the headset and the control box, which is connected by two cables.

The main disadvantages of the Pathsounder are:

- It is not waterproof; it will not work if is exposed to heavy rain.
- When the system will be dried is when it can be restored to normal function.
- The batteries loose strength rapidly at very low temperature.

The device has been tested during years with children. During the testing periods, the device gave confidence to the children. The device is considered an excellent electronic travel device for beginning orientation and mobility activities.

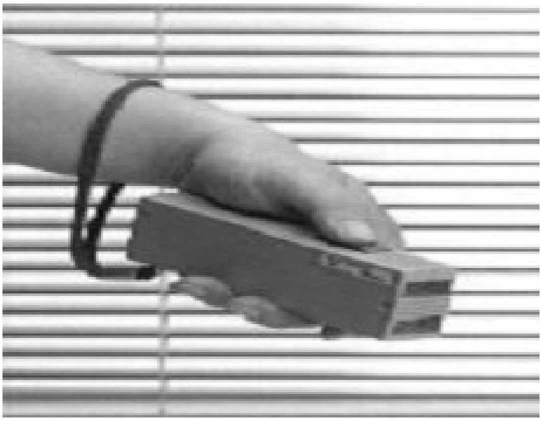


Figure 1.4 Mowat Sonar Sensor

Other ultrasound ETA system is the ***Mowat Sonar Sensor*** (Figure 1.4). It was developed by G.C. Mowat in New Zealand. The device is a hand-held system with sonar and tactile output. Mowat sonar is one of the simplest and most popular sonar sensor considered in the 80's (Morrissette, 1981). The device emits an elliptical ultrasonic cone of 15º width and 30º height, approximately, a human body form. The Mowat sensor detects only the nearest objects. When an object is detected, the sensor vibrates at a rate inverse to the distance from the object. The distance range is from 1 to 4m. There is a single control; three position slide switch on the top of the unit enable the user to operate the Sensor at two ranges. The longer range of 4m may be selected by pushing the slide switch forward from the off center position. The shorter range of 1m is attained by moving the control backward from the center position.

Like Lindsay Russell Pathsouder, the Mowat Sensor system is designed to be used as an additional informational system complementing the long cane or dog-guide both for indoor and for outdoor situations, to locate bus-stop signs, benches, doorways, other landmarks and pedestrian. Because of its small size, the sensor can be placed in the pocket without difficulties when it is not used (Freiberger, 1974), (ETA book, 1986).

Light Probes – represent simple instruments which are composed by photocells connected to electronic circuits with one or more outputs. These devices have the function of detecting through light probes the light source and transform it into acoustic, speech, tactile or vibratory outputs (Farmer, 1978).

The ***Nottingham Obstacle Detector*** – NOD (see Figure 1.5) is a small hand-held ultrasonic device, similar to Mowat Sonar Sensor, except on that it provides pulses of high frequency sound of 40 kHz (Farmer,1978). The pulses are eight notes, similar to a major musical scale (Freiberger, 1974). Like the Mowat Sensor, the device is useful for specific situations.

Figure 1.5 Nottingham Obstacle Detector (NOD)

The device has eight outputs, which represent the input data. The distance is broken down into seven small distances. A single note is attributed to each distance. When no objects are present in the area of vision of the sensor, no sounds will be provided. When no objects are in that range, the device is silent. When there is more than one sound, then the system will sonify only the nearest object. The maximal range of detection is 7 feet (2,1m) subdivided into eight 12 inches (30 cm) zones, each with its own individual signal tone.

The NOD aid is mostly used for obstacle localization in difficult situations. The aid has an on-off control and a small loudspeaker to supply audio output. The device has been researched for a tactile coding data representation (Amstrong, 1976).

Laser Cane – Light Amplification by Stimulated Emission of Radiation (see Figure 1.6) is a combination of the cane and the laser technology (Farmer, 1978) (Ando, 2003). Technical advances in miniaturized components such as integrated circuits, smaller batteries and especially intense solid-state, room temperature, gallium arsenide lasers made very compact housing possible.

Figure 1.6 Laser cane

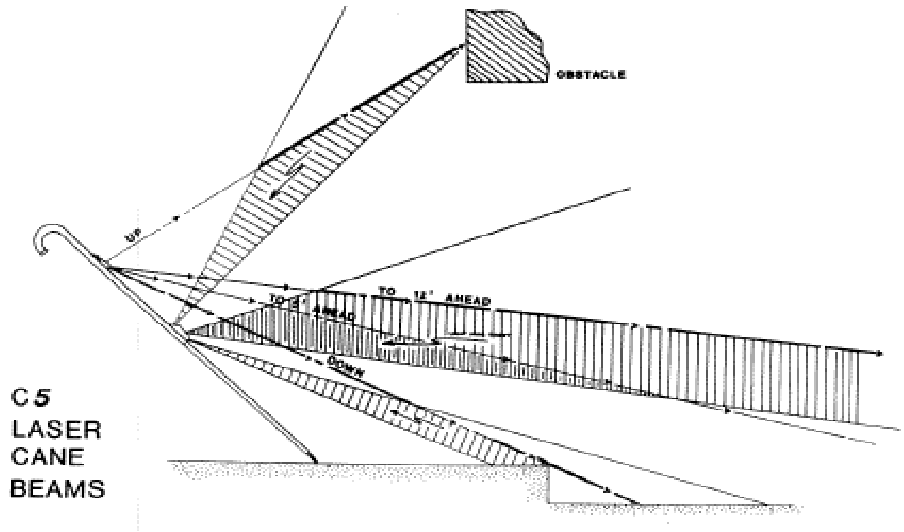


Figure 1.7 C-5 Laser Cane principle of object detection

In 1963 the C-3 Laser Cane became the first real Laser Cane, being evaluated by 50 blind users. Afterwards, an improved C-4 Laser Cane was developed. In 1971 and 1972, eight blind veterans were selected to participate in the trainings. The candidates were above-average travelers with travel experiences of at least one year beyond completion of their basic mobility training course.

The training lasted five weeks; then, each veteran took a C-4 Laser Cane, in order to test it at home. The C-4 Laser Cane evaluation results constituted the basis for the development of the improved C-5 Laser Cane, which was less bulky. The C-5 Laser Cane detects the obstacles in the range of 4m ahead. The system scans the environment and by acoustical signals informs the user about the nearest obstacle (see Figure 2.7). The sound is proportional to the distance between the obstacle and the user. The Laser Cane has been developed and manufactured by J. Malvern Benjamin and his colleagues from the Bionic Instruments Inc., under the name of Light Amplification by Stimulated Emission of Radiation Cane (Laser Cane). Laser Cane is a product of the combined effort between the private enterprises and the government.

The use of the Laser Cane by following detection information of the laser beam is based on the cane being used in the traditional touch techniques fashion.

The C-5 Laser Cane has three miniatures solid-state gallium arsenide (GaAs) room-temperature injection lasers which emit 0,2µs (micro seconds) pulses of 9050 Å (Angstrom), 40 or 80 times per second and three photosensitive receivers. In that case, the objects can be located with a high degree of accuracy by discrete scanning.

As it can be seen in Figure 1.7 the operation of the C-5 Laser Cane is based on the Granberg principle of optical triangulation applied to obstacle detection. A light beam is emitted by the sensor. When the light collides with the obstacle surface, it is reflected and received by the lenses of the photo detectors implemented on the laser system. In order to discriminate the range, the system is designed so that objects at different distances reflect lights on different receivers.

Sonic Pathfinder is an ETA system similar to Mowat Sonar Sensor and Sonic Torch, which detects the objects by receiving the reflected ultrasound transmitted by the device and informs the blind user via audio signals about the object presence (Debnath et al. 2004). The system emits tonal progression sounds through earphones when the user gets near the object. Unlike Mowat Sensor, Sonic Pathfinder is mounted on the user head, thus, the user hands are free. Nowadays, the similar devices are widely researched and improved with new technology (Debnath et. Al, 2004)

Figure 1.8 The Polaron device

Polaron ***TM*** is another ultrasonic ETA system within the category of clear path indicators or obstacle detectors (Farmer & Smith 1997). Polaron TM (see Figure 1.8) system detects the objects through ultrasonic sounds. Via vibrotactile or auditory signals, the device informs about the object presence in an area of 16 feet. The aid is designed to be used as a secondary aid together with the standard long cane or with a guide dog.

In these categories includes also the ***Sensory 6***. These systems detect the objects that are farther away than a distance of a long cane, using tones to indicate the user about object distance.

WalkMate also takes part of this category of object detectors which emit tones.

Among the most recent ultrasound ETA systems are the ***Ultracane***, a modern model of Batcane (Hoyle, 2003), ***Miniguide, Bat K Sonar Cane, KASPA, Trisensor, Sonicguide***, etc... These devices constitute the second class: *environmental sensors*, which try to get further than the mere detection of obstacles (ETA book, 1986), (Ando, 2007). These systems beside the object detection, display information about object texture and flow patterns. The most known one is the Kay *Sonic Torch*, developed in 1959 (see Figure 1.9) (Farmer, 1978), (Kay, 1964). The Sonic Torch is a hand-held narrow-beam ultrasonic ETA system which provides auditory signals as outputs via headphones. The pitch corresponds to range and the timbre corresponds to variations in target surface texture. Unlike the ultrasonic ETA systems within the class of object detection, Sonic Torch transmits a wide band-width frequency modulated ultrasonic energy wave (40-80kH).

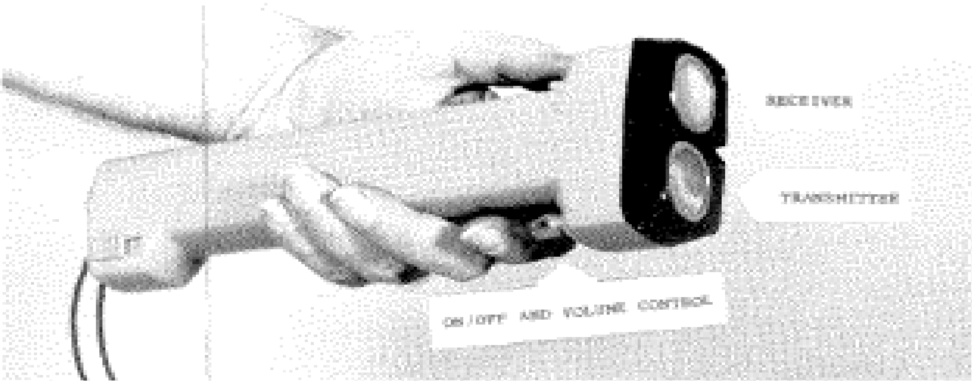


Figure 1.9 Sonic Torch

The Sonic Torch explores the area of view with a wide sonic cone of approximately 30º on either side of the midline direction in which was pointed. The system has been evaluated in more countries than any other sensory aid. Despite its wide evaluation, the system was unsuccessful in the sensory aids market due to its introduction and use as a primary travel aid and because it was a hand-held aid, which is uncomfortable for the blind users.

Binaural Sonar Electronic Travel Aid also known as ***Sonicguide*** (Figure 1.10), is the improved model of the Sonic Torch, one of the most influential electronic mobility aids built to date. Unlike the Sonic Torch (in 1966, Kay added to the Sonic Torch an additional channel, giving it the name of binaural), the Sonicguide is a head-mounted sensory aid. The Sonicguide was developed by Kay at the University of Canterbury, Christchurch in New Zealand by the Wormald International Sensory Aids Limited Christchurch. Telesensory System Inc. handles assembly and distribution in the United States of America, Canada and Brasil. The aid was designed to give the blind user more information about the surrounding environment (4m) through sounds, tactile or vibrotactile (Kuc, 2002). The sounds enable the blind user to estimate the distance between himself and the object by relating the pitch (Farmer & Smith, 1997).

The Sonicguide gives text instead of the headlines delivered by "go-no-go" devices, such as the Pathsounder and Laser Cane, as Russell suggests. Sonicguide supplies the user with three types of information: distance estimation; azimuth and directional appreciation and, finally, interpretation of tonal characteristics which make possible the object identification. Also, it provides rich information of the environment, only with practice and experience. The system has a sonic cone of 45º relative to the right and left of the body center. One of the main disadvantages of the system is that does not provide information about down steps and very low objects which appear on the travel path.

Sonicguide MkII is the secondary ETA system which requires conjunction with long cane or guide dog. The Sonicguide has maximum accuracy of 20ft (6m) with specular targets such as large, smooth surfaces like walls plate glace and an effective range of 12 to 15ft (3,6 to 4,5m) with diffuse objects (smaller or rough surfaces like trees or foliage).

Figure 1.10 Sonicguide Mk II glasses

During these previous decades the obstacle avoidance ETA systems evolve towards the *orientation aids,* which involve two aspects, the familiarization with the environment and the navigation (Brabyn, 1982, 1985). This class includes two approaches: modify the environment with electronic location identifiers or provide the traveler with an electronic device that locates him within a global or local coordinate system. The development of this type of ETA systems for navigation and orientation has a long history (Brabyn, 1997). Their development started at the Smith-Kettlewell Institute at the Georgia Institute of Technology. At first, they were based on used two main systems: the sensory system and video cameras.

Talking Signs System (Figure 1.11) is an infrared wireless communications system based on Remote Infrared Audible Signage (RIAS) that provides remote directional human voice messages that make confident and independent travel possible for vision impaired and print-handicapped individuals (Brabyn, 1982). The technology was pioneered and developed at Smith-Kettlewell Eye Research Institute, Rehabilitation Engineering Research Center in San Francisco, California in 1979. The infrared beam pattern provides control of range and coverage, and the directional nature of infrared light allows the user to accurately locate each sign. The clarity of the message increases with the direct pointing of the system to the object. The system works both in outdoor and indoor environments. Infrared transmitters are installed through environment such as in shopping centers, hotels, bus stops, airports or transport terminals.

Figure 1.11 Talking Signs device

Infrared transmitters are installed near the main environmental entities. Each transmitter continuously sends digitally encoded utterance about the nearby entity. The user having an RIAS receiver hears the utterance when pointing the receiver to the transmitter direction (see Figure 1.12).

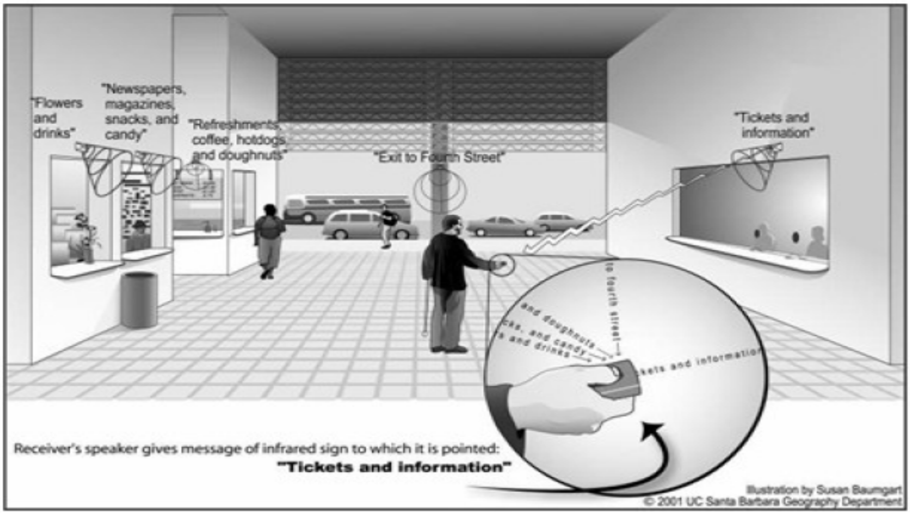


Figure 1.12 Talking Signs system function

The infrared radiation from the transmitters carried encoded information including speech about the identity of the location to which the transmitter is attached. The infrared radiation signal from the transmitter is highly directional, where the transmitter serves as a beacon. The system range is of 30m outdoor and lower for indoor environment.

SONA – Sonic Orientation Navigation Aid System relies on a similar idea. SONA system has been developed by the Georgia Institute of Technology after 1979. SONA is a prototype environmental labeling system with radio-activated auditory beacons to complete the navigation task (Kelly, 1981).

Marco is an infrared light transmitter and receiver system. Marco provides similar benefits for navigation as Talking Signs System. Its difference is that it has a range of 12,5m.

Verbal Landmarks is an orientation and navigation system which uses inductive loop radio signals instead of infrared lights (Loomis, 2006). The device has been developed on 1992, but is no longer in use. This device has a range of signal transmission of 2m. The system via spoken message informs the user about the obstacles, when he directs the system towards the obstacles. Device differs from Talking Signs in that the radiofrequency transmitter has a very limited range and it is omnidirectional (every direction), hindering the precise localization of the objects (Benzen and Mitchell, 1993).

Easy Walker is an innovative navigation system based on infrared device which provides the user the correct path to the selected destination (Ando, 2003).

Pilot Light also known as mini-radar is an infrared hand-held navigation device with an adjustable range up to 3,5m. The system is responsive for infrared talking signal.

Some of the devices of *orientation and navigation aids* are GPS based systems. This group embraces the navigation aids based on Global Positioning Systems (GPS), in which satellites continuously send signals to receivers (Farmer, 1997; Collins, 1985; Loomis, 1985).

The idea of using GPS systems to assist navigation to the visually impaired people comes back around two decades (Collins, 1985; Loomis, 1985). Thanks to the Global Positioning satellites which provide a tool for labeling every point of the Earth surface using longitudinal and lateral coordinates, nowadays GPS receivers can provide in real time the information about coordinates of the user and manage geographical data base for getting maps and landmark details. Currently, GPS systems are able to provide user constant updates about their location, including streets they are walking on, the travel direction, traveler speed, building identification and lightings of intersections.

These systems provide a wide variety of orientation and navigation support to the blind user. They inform the user about the route by verbal, waypoint, landmark, route instruction and other comments, in a range of 20m. During the last decades, many navigation systems have been developed based on compass, GPS or DGPS.

Pedestrian use of GPS for positioning has three shortcomings. One is that the accuracy of stand-alone commercially available GPS receivers is limited to 20m. A higher accuracy is achieved by the differential correction DGPS, in which the correction signals from GPS receivers, are transmitted by radio link to the mobile receiver, allowing later to determine the position with an absolute positional accuracy on the order of 1m or better. However, the DGPS have their disadvantages, since they require a separate receiver and the service does not cover all locations. The second outcome is the satellite connection, which has loss of visibility when nearby buildings or dense foliage block a substantial part of the sky. The third shortcoming involves multipath distortions resulting from reflections of the GPS signal from nearby structures, because the satellite considers the time delay of signal transmission and signal perception, so positions derived from reflected signals constitute an error.

Among the main previously mentioned systems we can find the Mobility of Blind and Elderly People Interacting with Computers (***MoBIC***) project (Petrie et al, 1997), Makino, Electronic Guide Dog, GPS Braille Note, Trekker, Geotact etc.

The MoBIC aid consists of two integrated systems (Strothotte et al, 1998): the MoBIC Pre-Journey System (MoPS), which plans activities, and the MoBIC Outdoor System (MoODS), which is in charge of plans executions, to guide the blind user during journeys (Douglas, 1997). The user, through a small hand-held keyboard asks the question to the system and he receives the answer acoustically (Duglas et al., 1997).

Makino is an ETA system which uses a digital mobile phone, which enables the communication between the user and the computer database. The mobile phone transmits the GPS coordinates to the computer central unit and it provides the position to the user via synthetic speech (Loomis et al., 2001).

Electronic Guide Dog has a technology similar to Makino, i.e, a mobile phone and a central unit. Unlike Makino, Electronic Guide Dog uses human agent at the central facility, who communicates via voice with the user giving him the positions and other information (Loomis et al., 2001).

Personal Guidance System (PGS) is a GPS based system for navigation and orientation (Loomis et al, 2001). The goal of designing the PGS was to be portable and self-contained system which would allow visually impaired people to travel through familiar and unfamiliar environments without guide assistance (Loomis, 2006). The system is based on a Fluxgate Compass, digital maps (Loomis 1985) a GPS and computer (Looms et al, 2006, 1998). The GPS detects the user location, by connecting with a Geographic Information System (GIS) computer and, via earphones, the traveler hears synthetic speech (Loomis et al, 2005), or via vibrotactile (Klatzky, 2006) gets informed on his location. Despite its technology the system is designed to be used as a complementary system in combination with the long cane, seeing-eye dog or other ultrasonic navigation aids.

The original idea of the PGS system is that, when a visually impaired person travels through the environment, he/she can hear the names of the streets, building, etc., spoken by a speech synthesizer, coming from the appropriate locations in auditory space, as they come from loudspeakers at those locations. Figure 1.13 is a graphical representation of the early conception of the system.

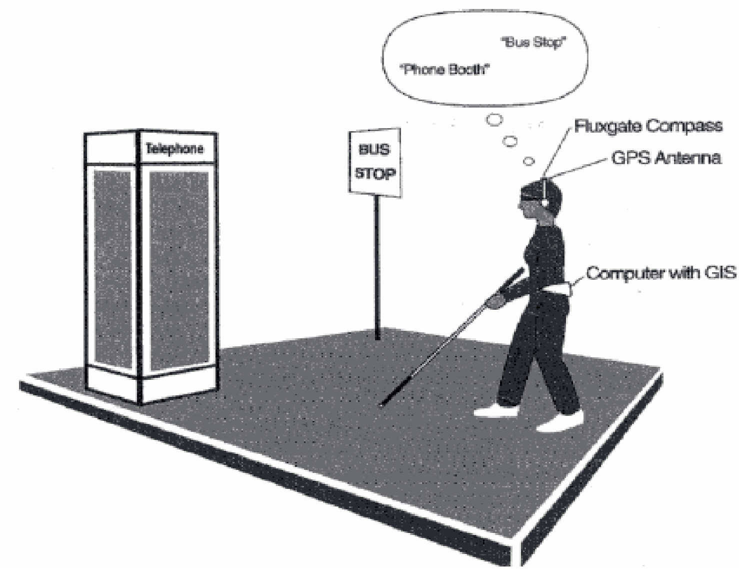


Figure 1.13 Early conception of the Personal Guidance System

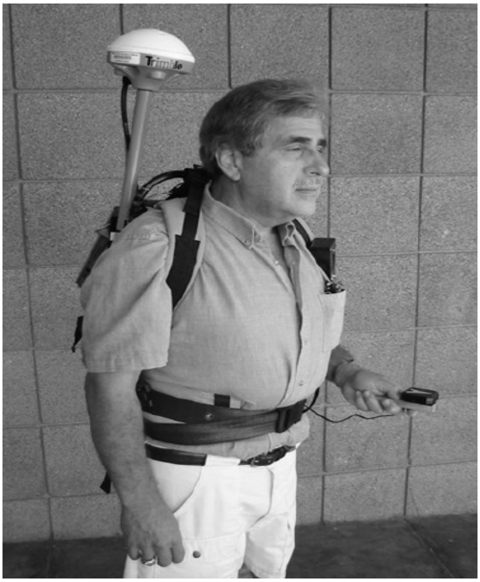


Figure 1.14 Haptic Pointer Interface of the Personal Guidance System

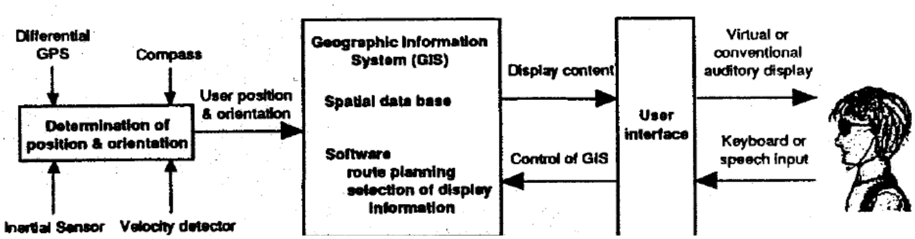


Figure 1.15 Functional components of the Personal Guidance system

PGS has been designed for the long distance, taking into consideration that, since it is a complementary system, the user via the long cane, guide-dog or ultrasonic sensing devices can detect the immediate environment.

Via input interface, the user introduces the destination into the system and controls the various modes of operations. The usual input interfaces for the commercial products are the Querty Keyboard and Braille Chording Keyboard.

The display interface for navigation system provides at least two types of information. The first type is the route guidance information, which guides the traveler from the origin to the destination. The second type of information is the off-route information, which can both help the user to keep oriented and develop better mental representation of the surrounding environment over multiple trips through it. The system range is 20m (Loomis and Golledge, 1993).

An important function of the system is that the user can learn about points of interest in the environment such as restaurants, stops, etc… contributing to the construction of a cognitive representation of the spatial layout of the environment.

GPS Braille Note is based on Braille Note system; a GPS receiver, maps and points of interest databases that provide spoken and/or Braille access to location information in any outdoor environment. GPS Braille Note aid was created by Sendero Group. With that system, the user can explore an unknown area, follow routes, select destinations, etc… (Humanware).

Figure 1.16 GPS Braille Note

Geotact system is a GPS sensor-based system with an inertial sensor to calculate and reduce the effects of degradation of GPS (Fancy et al, 2006). Unlike other GPS navigation systems, Geotact provides speech information on distance in meters and direction in time, coordinates as “2 o’clock, 160 meters for right, nine o’clock for left” etc. With Geotact system the user is able to select the way for traveling towards his objective, because the system does not decide the route to go. The system only gives possible trajectories, among which the blind user must select his favorite.

Easy Walk system works with a GPS satellite receiver and a mobile phone with Symbian operating system. The modern aids include the Bluetooth GPS receiver (Action for Blind people). 24 hours call-center is responsible for requesting and informing the user always he requires their service. The system is also equipped with a service called Easy Contact, which allows deaf and mute people, to send SMS to the call-center in order to call a taxi, to book a room in a hotel or to make a reservation in a restaurant.

Figure 1.17 Easy Walk system for navigation

Talk & Wayfinder which includes a mobile phone, Bluetooth GPS receiver, has become a guide for European and American blind users. Thanks to the Talks software, the system provides access to the mobile phone functions by using a voice-led menu and vocal mode. The GPS system helps blind users to plan complete routes from one point to another by means of vocal output. Like a car GPS, the Talk & Wayfinder enters the destination and the system calculates the itinerary.

TORMES a hand-held device based on the EGNOS technology, GPS receiver and Braille Keyboard enables visual impaired people in their navigation through known and unknown environments (Morales and Berrocal, 2005). Via a synthesized speech, the user receives complete information about his location and directions. The system has an accuracy of 2m, (the GPS has 15 to 20m of accuracy). The device has been developed by the ESA Company with ONCE and European Commission funding. (CIDAT), (ESA, 2003), (ESA, 2008).

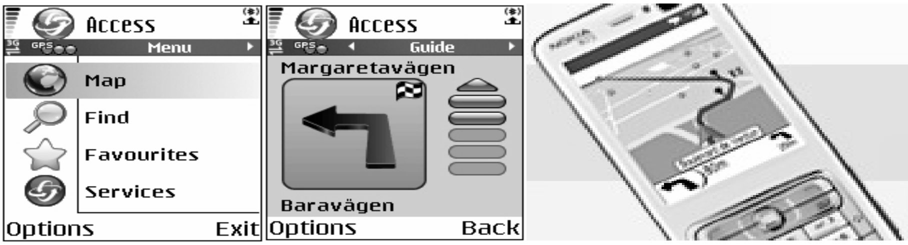


Figure 1.18 Talk & Wayfinder

Figure 1.19 Tormes device

Loadstone GPS- navigation aid for blind is a navigation solution for the blind and visually impaired community. Loadstone GPS is similar to the Wayfinder and Trekker systems, although the system uses the well defined "points" instead of using maps like the previous aids (www.loadstone-gps.com). The device is based on a mobile phone and Symbian operating system with Bluetooth GPS receiver, USB Bluetooth dongle, Bluetooth headset, Multimedia card reader (MMC). Thanks to its operating system, Loadstone GPS is also called as “Open Source”. Via voice, the system informs the user about routes marking them with points (Forum Nokia), (The Loadstone GPS team, 2008).

Trekker is a GPS system for the blind and visually impaired which uses a GPS and digital maps in order to extract the information on the user location and select destinations with high precision. Tekker is designed to be used as a complementary aid, complementing the existing aids, white cane or guide dogs, but not replacing them. The maps can be downloaded and saved in the aid. The system provides real time information, maps and object detection. Via vocal speech, the system informs the user about his location and provides other information. It is flexible and it plans routes and records them (Lagace, 2005).

Figure 1.20 Trekker a GPS system for the blind and visually impaired

1.3 Navigation Projects

Blindness is more feared by the public than any ailment, except on cancer and AIDS. During many years, the researchers have tried to substitute the human vision by artificial vision. Despite intensive efforts, the restoration of fully functional vision in blind people has not been achieved. The human brain has a very complex functionality model, and the scientists does not know enough about how to communicate with the altered cortex in order to generate meaningful visual perception. Other class of *navigation aids* includes the navigation systems based on *artificial vision*. During these previous decades, different researchers have been working hardly on human vision substitution by artificial vision, on determining the basic parameters for visual prostheses in order to restore the useful reading abilities (Fornoz et al., 2005). Despite all efforts, it was difficult to learn about how to communicate with the altered cortex in order to generate meaningful visual perception (Amedi et al., 2004).

One of the pioneer works carried out on object detection using artificial vision is the work developed by Bach-y-Rita (1969) in which a television camera and a 20×20 array of four hundred solenoid stimulators was used. The system delivers the object shape perceived by the camera to the array of stimulators built into dental chair. It reproduces the shape of the detected object through applying corresponding vibrations.

***TANIA* - a Tactile Acoustical Navigation and Information Assistance** is a hand-held aid based on maps, a sensory system and tactile-acoustical interfaces (Hub, 2007). The system is based on a portable tablet PC, an MTx inertial sensor, and a Braille display. The aid can be used to search tasks within the tree-dimensional environment (Hub, 2004). The system can recognize the room numbers, distance of the route etc. Via acoustical sound and on the torch screen of the tablet PC, the information perceived by the system is visually represented, by pressing the pushbutton (Hub, 2008).

Figure 1.21. TANIA navigation aid

SWAN **– System for Wearable Audio Navigation** is a wearable computer consisting of a tactile input and audio output which guarantees the navigation and object detection.

Figure 1.22 The user with the SWAN system

The system uses the Geographic Information System (GIS) in order to spatialize the data (Wilson et al., 2007) and the spatialized non-speech beacons guide the listener along a predetermined path, from a starting point, through several waypoints, and arriving at the listener destination (Walker, 2005 and 2006). It is based on four cameras, light sensors, GPS, digital compass and a laptop. Swan system determines what features in the environment need to be presented to the listener and renders the appropriate sounds via stereo headphones.

The environmental elements that are sonified in SWAN include navigation waypoints: a variety of surface transitions such as carpet to tile, corridor level, stairs, a variety of objects like offices, benches, mailboxes, etc. The beacon sound represent the waypoint location along the path.

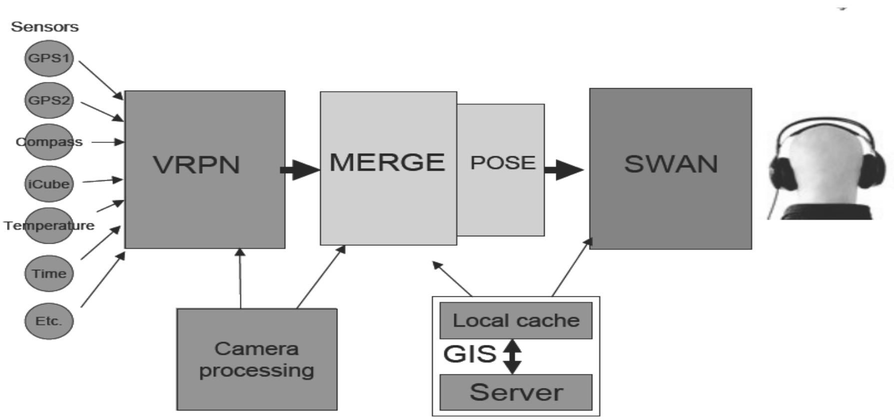


Figure 1.23 SWAN architecture

EPFL project is a project which aims to develop a device to detect obstacles on shoulder height via a stereoscopic sonar system and to send back a vibrotactile feedback to inform the user about his localization. The project is coordinated by Cardin Thalmann and Vexo from Ecole Polytecnique Federale de Lausanne. The project consists of sonar sensors, a microcontroller, 8 vibrators and a calibration console.

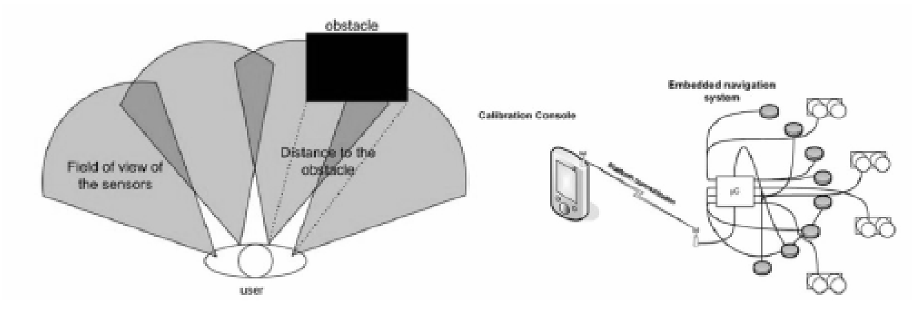


Figure 1.24 EPFL prototype design and operation area

The advantages of this project are: It is a wearable light, low power consumption and low-cost system.

A new complex navigation system based on synthetic vision, an open interface for object, colours and face recognition, auditory and tactile display has been developed within ***The vOICe* system Seeing with sounds** (Meijer, 2005). vOICe allows to represent visual information with sounds. A camera scans the visual fields. The computer converts video images acquired by the camera into sounds (Meijer, 1992) or tactile maps. The scene in front of the user is scanned in stereo; this means that the objects from the left will be listened on the left by the left ear and the objects on the right will be heart on the right by the right ear. It also shows how the sounds might "look": Brightness represented by volume; elevation is represented by pitch.

Figure 1.25 The vOICe system components

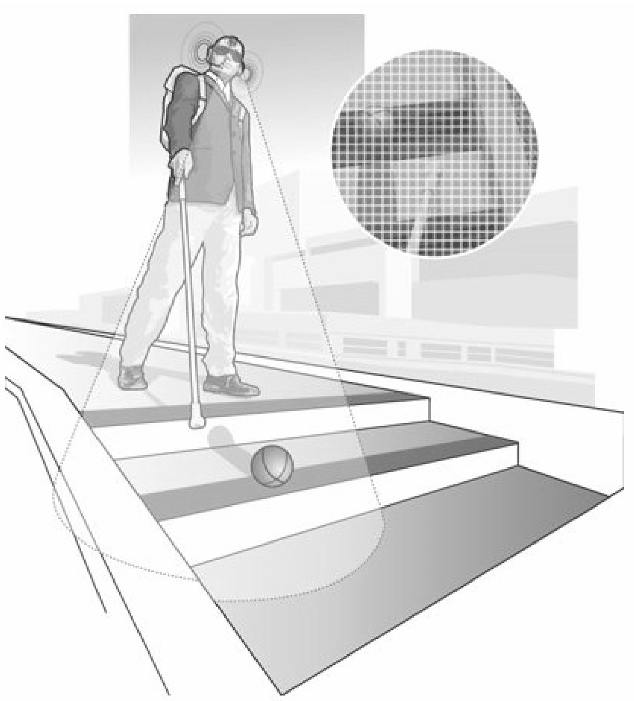


Figure 1.26 vOICe functionality methodology

Figure 1.27 represents the last version of vOICe device; is is implemented on a 3D webcam Vizix Wrap 920AR HMD model with USB powered stereo camera pairs.

Figure 1.27 Last version of vOICe device

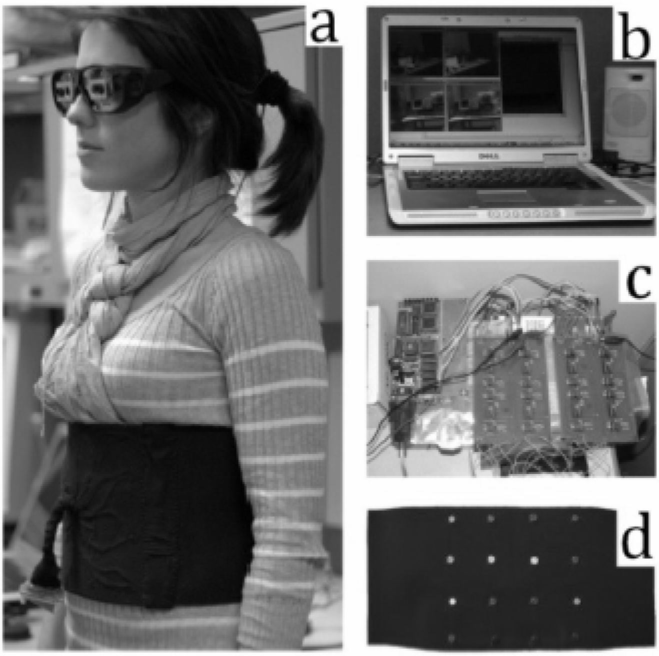


Figure 1.28 Tyflos prototype 2nd version components. a) Stereo cameras attached on dark eyeglasses and vibration array vest on the user abdomen, b) portable computer, c) microcontroller and PCBs, d) arrangement of the 4x4 vibrating elements inside the vibration array vest

One of the most important researches on navigation systems using artificial vision is the ***Tyflos device*** (Bourbakis and Kavraki, 1996) which was initiated in mid 90's. Tyflos was designed to complete two functions: reading and navigation (Dakopolous and Bourbakis, 2007). Tyflos Navigator captures the information from the surrounding environment via two tiny cameras, a range sensor mounted at the branches of a pair of black glasses, a GPS device and a RFID reader (Bourbakis, 2003). Via a speech-recognition and 2-D vibration interface, the user interacts with the system (Bourbakis and kakumanu, 2008), (Dakopolous and Bourbakis, 2008). The main goal of the Tyflos system is to integrate different navigation assistive technologies such as a wireless hand-held computer, cameras, range sensors, GPS sensors, microphone, natural language processor, text-to-speech device and a digital audio recorder. Moreover, it includes methodologies such as recognition based segmentation, range data conversation, fusion etc., in order to offer the blind user more independence during navigation and reading.

The main advantage of the Tyflos is that it is free-ears and that the use of the 2D vibrations array with the variables offers the user a more accurate representation of the 3D environment giving also information for distances.

EyePlusPlus or the **Forehead Sensory Recognition System (FSRS)** is an electronic travel aid that should be used by appropriately trained individuals to complement existing travel aids (e.g., white cane, guide dog, etc.) by providing additional sensory information via electrical stimulation regarding the user's physical environment. The FSRS converts video images captured by a miniature sunglasses-mounted camera into electro-tactile stimulation patterns applied to the user's forehead through an array of stimulation electrodes, which allows the user to perceive objects.

Figure 1.29 Forehead Sensory Recognition System components

NAVI Navigation Assistance for Visually Impaired – is a sound based obstacle identification device for blind people. The system has been developed at University Malaysia Sabah by Sainarayanan and his colleagues (Sainarayanan et al, 2007). The system is composed by one digital video camera, stereo headphones, SBPS (Single Board Processing System) with chassis, NAVI vest and recharged batteries. The system detects the objects in front of view, a 32x32 resolution gray-scale video is captured by the camera. The SBPS uses a Fuzzy LVQ neural network in order to classify the pixels transforming the grey level features into stereo sounds. With the system, the users were able to detect static and slow moving objects. A great disadvantage of the system is that it does not detect the distance of the object.

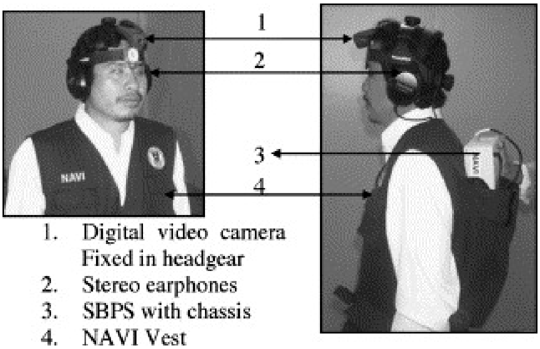


Figure 1.30 NAVI system

1.4 Human stereo vision or stereoscopic vision

The "stereo" comes from Greek "stereos", which means *firm* or *solid*. Humans and all animals, insects etc. have two eyes side-by-side in the front of the head. Separation of the eyes enables taking a view of the same scene from slightly different angles. The brain, simultaneously receives this two images combining them into a single image. Figure 1.31 represents human stereo vision. Beside the combinations of two scenes, stereovision is able to detect colours, the object placement (the difference between the object and the body in distance, azimuth, etc.,) to perceive an empty space (the space which is between the objects) etc.

Human stereo visions have the next components:

- convergence of the eyes (eyes muscle) until double vision is overcome
- focuses of the lenses to render object sharp
- physiological diplopia (3D vision depends on repeated corrections for double vision)

Stereo vision technologies have the same process as the human stereo vision, capturing the environment information with two cameras processing and representing this information in a useful way.

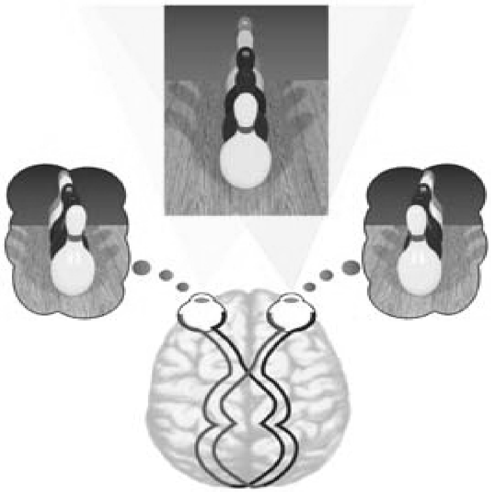


Figure 1.31 Human stereo vision

1.5 Conclusions

In the present chapter forty one commercial and prototyped devices are described and classified by their principle of work and components: obstacle detectors, environmental sensors and navigation. This study show that none of the devices produced up to date achieved all criteria for successful mobility for blind people: safety, comfort, grace and independence. In part this has been due to the limitations of the technology. From this limitation takes part the information limitations with mean that the developed devices are capable of completing a partial part of the criteria mentioned below, training difficulties, relation to the white cane and guide dogs, and auditory cues and expense. The most important findings are that the visual impaired users are not confident with the device, because the most of the systems are at the prototype stage and/or need long experimental time with blind people. In summary the there need an electronic travel aids that are inexpensive – is necessary to be available for all social class of blind people; simple - easy to learn and use; wearable – to allow blind user to use advantages of wearable technologies; and not distracting from natural cues – there is necessary that to listen environmental should not be interfered, to show as much possible information of the environment; free-hand – is not required by the blind users to be hand-held device because they will hold the white cane.

2. CHAPTER II: ACOUSTICAL NAVIGATION SYSTEM FOR VISUAL IMPAIRED PEOPLE

CHAPTER II: ACOUSTICAL NAVIGATION SYSTEM FOR VISUAL IMPAIRED PEOPLE

2.0. Summary

In the present chapter, a portable prototype capable to transform 3D environmental information into acoustic stimuli is described. Unlike other devices mentioned in the chapter 1, which are based on sensory acquisition systems, the proposed device consists of a complete working prototype which combines the sensory, artificial vision and GPS acquisition system and speech and acoustic output system. Implementation of these three technologies for the environmental data acquisition enables the system to improve its working area and its quality. Due to the three acquisition technologies, the system integrates the object detection and orientation and navigation Electronic Travel Aids into one single unit, being able to work in an autonomous way or altogether. The cognitive part off the device lies in the fact that it implements a methodology for simulating that a series of sounds, in a virtual way, are radiated by the user surrounding objects. These sounds are capable to carry accurate spatial information. The main goal is to generate in the user a correct perception of virtual sound sources emitting from the object surfaces, which aims to allow the brain to create a three-dimensional perceptual image of those objects like in the real world. Using these ideas, it is aimed to create a global perception of the sound, enabling visual impaired people, on real-time, to perceive and get a global image of the surrounding environment and the way the objects are organized.

The chapter is distributed in five sections. The first section describes the system design and requirements of the device. It includes the characteristics of the system and the technical requirements for the system development. During the second section, the sensory system is described. It includes the hardware implementation and perception range and characteristics. Unlike previous sensory or laser systems, the developed device is able to detect and process the information in real time. In the third section, the development of the artificial vision system is described. Also the object detection, camera calibration, image segmentation and depth map estimation methods are presented. In comparison with other methods, which use stereovision for object detection, the present device uses the methodology of the moving object detection for near and far distances. In the fourth section, we discuss the development of the acoustical system and its importance in the navigation systems for blind people. A description of the sound generation

and Head-Related Transfer Functions calculation and the sound validation is presented.

In the final section we discuss about the communication interface between the sensor, artificial vision and acoustic interface. The developed device described in the present chapter is later used for the experimental trials discussed in chapter five.

2.1 System design

Humans use a wide range of information for navigation such as depth, azimuth etc. Due to this fact, it is not easy to decide which information can be the best acoustically interpreted. Therefore, before building the system it is extremely necessary to define the aspects of the visual scene which represent the most important features for navigation and object identification (presence of the objects and their position in space). The auditory system, which is capable of combining information by classes of cues and by frequencies in order to synthesize a unitary spatial image, plays a crucial role for the navigation. Indeed, the auditory system solves a difficult problem when localizing sounds, mainly when there is more than one sound source (Takahashi T. T., Keller C. H., 1994).

The main drawback of the existing systems is the complexity of the computational algorithm and the high cost of the necessary resources. Regarding the speech navigational systems, they give precise information, although they can not provide real-time information. In addition, they can be confused with human speech.

In order to solve all these disadvantages, a cognitive system, which would provide access to the spatial information surrounding the user via stereovision system and audio interface, was developed Virtual Acoustic Space (VAS) (Mora J.L. G., Rodriguez-Hernandez A.F., Martin F., Castellano M.A., 2006).

Taking as bases the blind community necessities, the previously developed devices and the results obtained with Virtual Acoustic Space (VAS) developed at the Universidad de la Laguna, it appeared the idea of developing of a new simple and complex device which will help blind and visually impaired people to navigate through known and unknown environments. If such a system was developed, it would overcome most of the disadvantages of the existing obstacle detection, environmental sensors and navigation systems. On the other hand, it is necessary to develop a commercial system, available to all social classes, thus, cheap; a system which will be easy to be used and that will contain the maximum technology and the minimum complication on usage. The requirements necessary to implement the mentioned device are shortly described next.

2.2 Requirements for the system

Usage characteristics of the system:

- For blind and partial sighed people.
- Not to be heavy.
- To inform the user about the surrounding environment with minimum errors.
- Not to disturb and irritate the user.
- Not to hurt the user.
- Not to disturb the user to perceive the surrounding information by his eyes/ears/touch, etc.
- To be a free hand system.

Technical characteristics of the system:

- To use high technology.
- To be a cheap device.
- To be easy to be used.
- To have a clear and simple user interface.
- Real time environment acquisition system.
- Real-time processing system.
- Real time reproduction system.
- Small volume of the device.
- Accuracy.
- Compatibility between different technologies and sub technologies.
- To cover the nearer and farther areas of vision.

According to the requirements to implement the system, in Figure 2.1 and Figure 2.2 the main components of the navigation system are represented (input, processing and output components). The basic input components have been selected to complete the task for object detection, environmental sensors and orientation and navigation.

For object detection, a sensory system using 3D CMOS Time of Flight sensor and stereo-vision system using two FL2-03S2 cameras with auto-iris control has been developed. Each of these two systems is designed to work together and individually, depending to the user wishes.

For the orientation and navigation system, the use of GPS plays an important role, because in order to calculate the user position in both the stereo-vision system and in the GPS system, the HPS was used. Due to the HPS, the stereo-vision system calculates the moving object speed and it can determine its real position.

For the processing system, a Toshiba Satelite Pro notebook was used for the whole device processing. The sensor system uses a special technology developed on a FPGA. The GPS system works with a small PDA. The environment reproduction and cognitive loads are developed acoustically and vocally. Acoustical sounds inform the user about nearer and farther objects that appear in the area of vision of the both systems. The GPS system uses voice information to tell the user his position and the route he should follow.

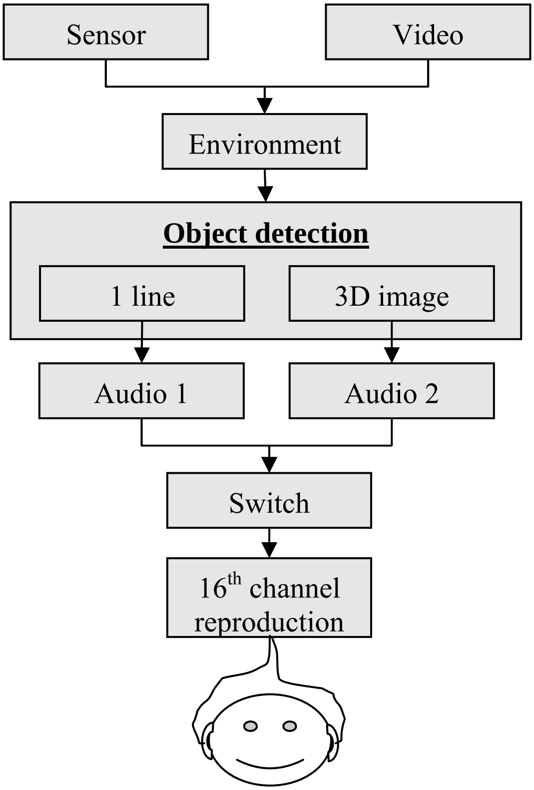


Figure 2.1 General working procedure of the system

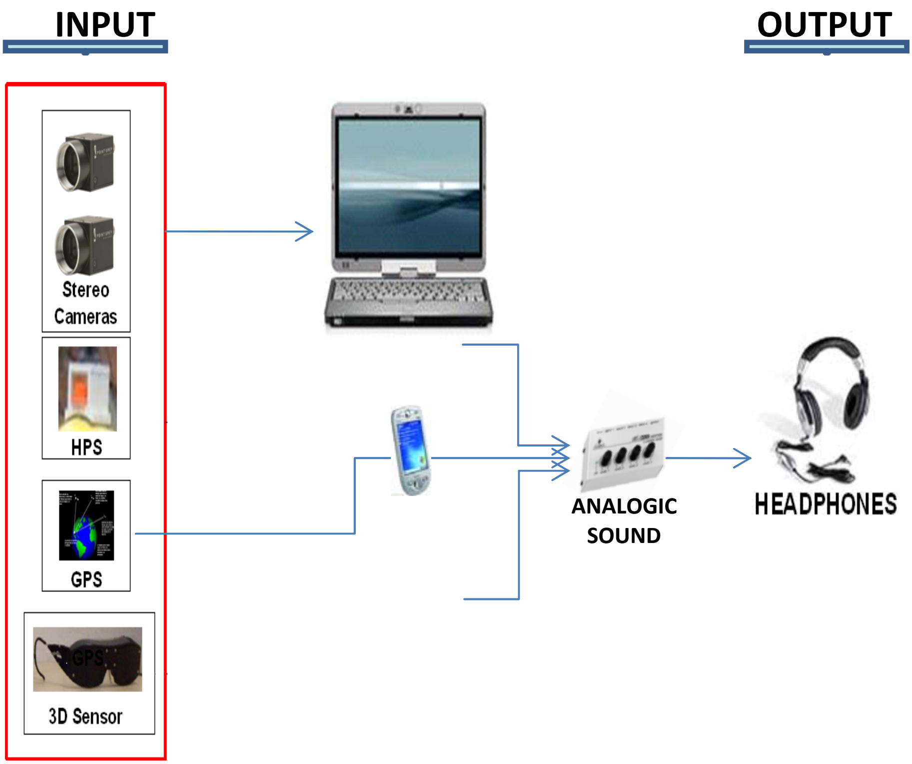


Figure 2.2 Technical representation of the system components

The output system is based on simple headphones, which are connected to the notebook.

2.3 Sensor System

It is desirable for the visual information input unit to be small and lightweight, because all these devices will be mounted on the user head. The sensory system with all the optical components, analogue and digital electronics and laser were assembled on a pair of glasses, as shown in Figure 2.3.a. The maximum distance reached by the sensor is 5 m at 64º in azimuth. The measurement principle is based on Time of F measurement of pulse-modulated laser light using a high-speed photosensitive CMOS sensor and infrared laser pulse illumination (Figure 2.4). The infrared impulses are emitted by the CMOS sensor. When any object appears in the front of the laser, the infrared impulses crash the object and send infrared impulses back to the system. When the impulses arrive to the system, it calculates the distance between itself and the object. The analogue signals of several laser pulses are averaged on a chip to reduce the required laser power and also to increase measurement accuracy. A fully solid state micro system is embedded on FPGA (Figure 2.3. a). The advantage of using these sensors is to provide an exact distance both in a horizontal and in a frontal plane. In addition, they reduce the necessary processing time for the calculation. The information from CMOS sensor is used in the audio representation module when decisions are taken for generating the appropriate sound map.

a)

Figure 2.3 Real system composition: a) Components of the operational system b) Full device picture.

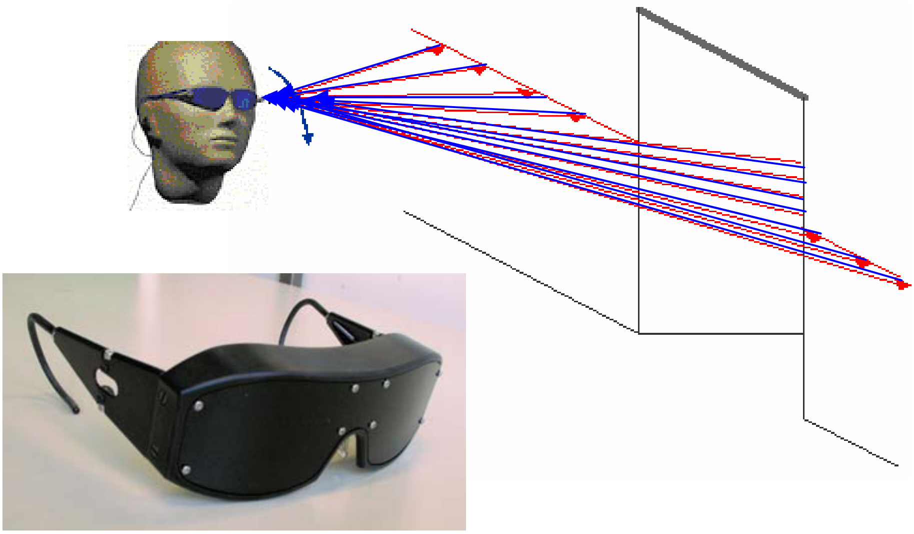


Figure 2.4 The working area of the 3D-CMOS sensor with laser illumination implemented into a pair of glasses. The Red colour represents the infrared impulses emitted by the CMOS sensor; the blue colour represents the laser pulse reflection from the object.

The system is based on a computational algorithm which assigns a two dimensional position to any object falling in the overlapping region of the sensor field of vision. From each of the 64 pixels of the sensor, a light beam is emitted. In this way, the exact distance between the sensor and the object for a specific azimuth angle is obtained.

The Computational Algorithm attributes to each pixel, according to the obtained distance, a spatialized random sequence of short sounds. It is important to remark that an important innovation of the proposed system is that two-dimensional environment information is acquired and represented by two-dimensional spatialized sounds. Figure 2.5 shows the general scheme of the developed system. That figure shows schematically the processing steps for converting the sensor data input into an acoustic stimulus.

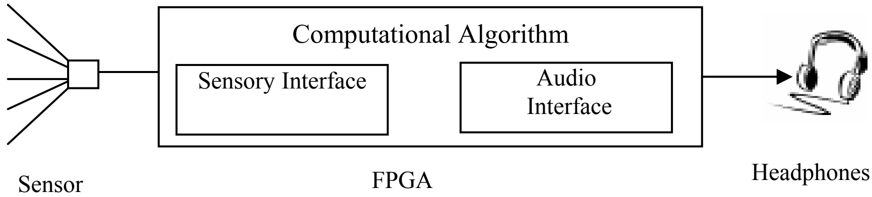


Figure 2.5 Schematic representation of the Sensory System

2.4 Artificial Vision System

A more complex and optimized device of CASBliP prototype is the Real–Time Assistance Prototype based on artificial vision.

The hardware of the Real-Time assistance Prototype, the Artificial Vision Subsystem is composed by the following main components:

1. A stereo camera system, preferably as small and light as possible, mounted on a normal lightweight pair of glasses.
2. Wearable PC for processing the software.

A description of the subsystem requirements and specifications is included in the Annexe III Camera Specifications

Figure 2.6 represents the actual device of the Real–Time Assistance Prototype, in which the two firewire CCD colour cameras (IEEE 1394b) are mounted on a helmet with a maximum resolution limit of 1024x768. Unlike the previous artificial vision, based on navigation systems, the Real–Time Assistance Prototype uses static and dynamic object/obstacle and free path recognition (Figure 2.1) and acoustical interface. The head mounted system is able to classify the obstacles according to their level of danger for the user, estimating their position, speed and direction of motion.

Two FL2-03S2 cameras with auto-iris control and a 6dof inertial sensor were used as main components of the vision system. A small Toshiba laptop was used as an operational system. Acoustically, the system warns the user about the objects within the area of vision in a perimeter between 5 and 15m in distance and 64º in azimuth. The Real–Time Assistance Prototype emits short acoustical signals through headphones at a rate of 64 pixels per image at 2 frames per second. A high resolution collection of non-individual filters (HRTFs) has been measured and synthesised with the set of sounds. When no objects are in the visual area of the system, there is no output. When one or more objects appear in the visual area, the system processing unit analyses the received information converting that data into acoustical signals. Then, the users listen via headphones, the sound corresponding to the objects position. The auditory output increases as the object is getting near the user. The implemented strategy is expected to generate in the user a visual-like auditory perception. According to the listened sounds, the user is able to take a decision in order to avoid the obstacle appearing in the system area of view. With a minimum practice, the user can easily detect the size of the obstacles. Different acoustical outputs were generated for obstacles and free path.

The vision subsystem processes the scene information, representing a limited area of the subject's frontal view defined by the stereo camera system's aperture, zoom and focus settings. Video acquisition utilities, camera control software and Visual C++ development environment is used to develop the software.

The software performs the scene segmentation tasks and generates the appropriate scene description signals that will be delivered to the next stage of the processing. It is fed from information supplied by the SIEMENS depth sensor to supplement the scene segmentation and coarse depth analysis information, it extracts from the stereo image sequence input.

The methods of scene segmentation and depth derivation are currently under investigation. A basic version for fast and simple object segmentation is initially aimed for before attempting to integrate information from the SIEMENS sensor. The methods developed take into account the scenarios set out by DBSV in their requirements document:

1. Crossing at traditional/complex intersections. However, it is important to note that any system developed for such a purpose will not automatically work at crossings in different cities, in different countries, due to extreme variations in environmental and physical conditions.
2. Navigation through known and unknown environment, streets, parks, gardens, commercial centres.

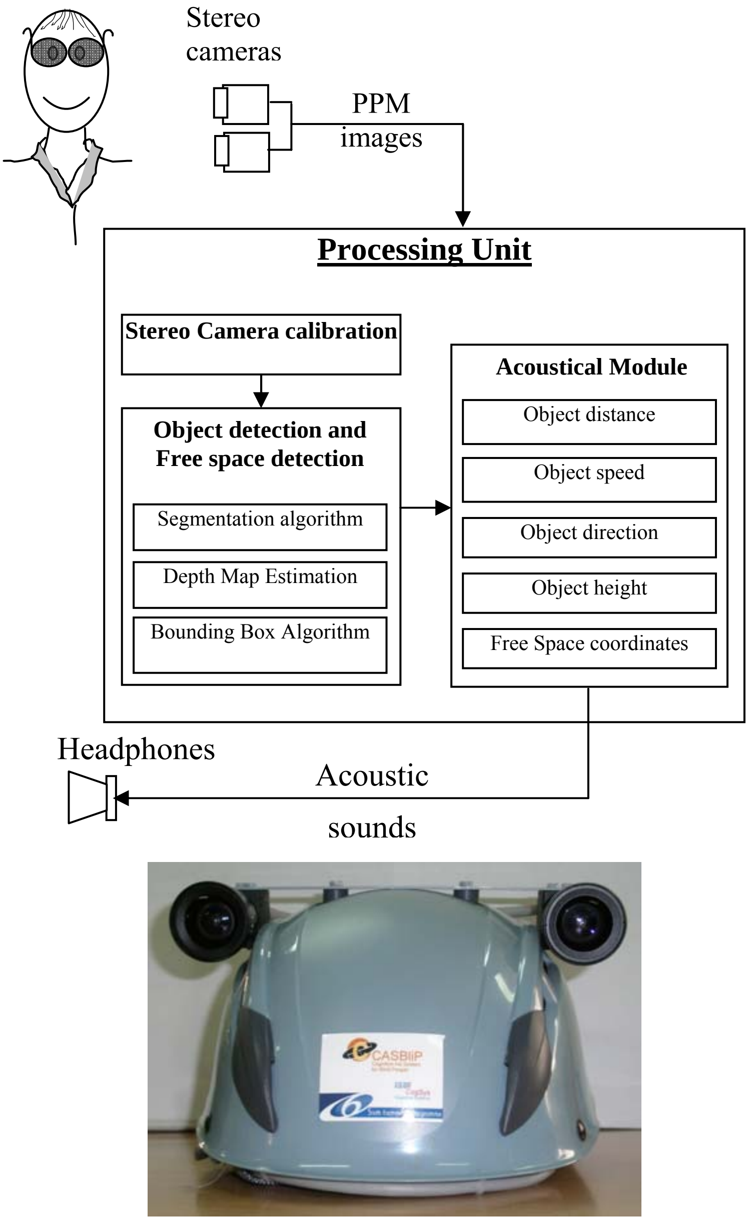


Figure 2.6 The picture of the Real–Time Assistance Prototype and the block diagram

2.4.1 Camera calibration

Specific self-calibration algorithms for restricted motions, combining camera motion constraints and camera intrinsic constraints were proposed by many researchers. Some interesting approaches were proposed by Moons et al. (1996) for pure translation, Hartley (1994) for pure rotations and by Armstrong et al. (1996) for planar motion.

A specific combination method between self-calibration and scene constraints was proposed in Pollefeis (1996) to resolve a case with minimal information. A method of elimination to impose the scene constraints was proposed by Bondyfalat and Bougnoux (1998). A combined approach was achieved by Liebowitz and Zisserman (1999). They formulated both the scene constraints and the self-calibration constraints as constraints on the absolute conic.

The problem of critical motion sequences is another important aspect of the self-calibration problem.

The GUI was used for images synchronization. Figure 2.7 show the used software for stereo camera calibration.

The checkerboard was hand labelled with features and motion of these featured reconciled with the sensor output.

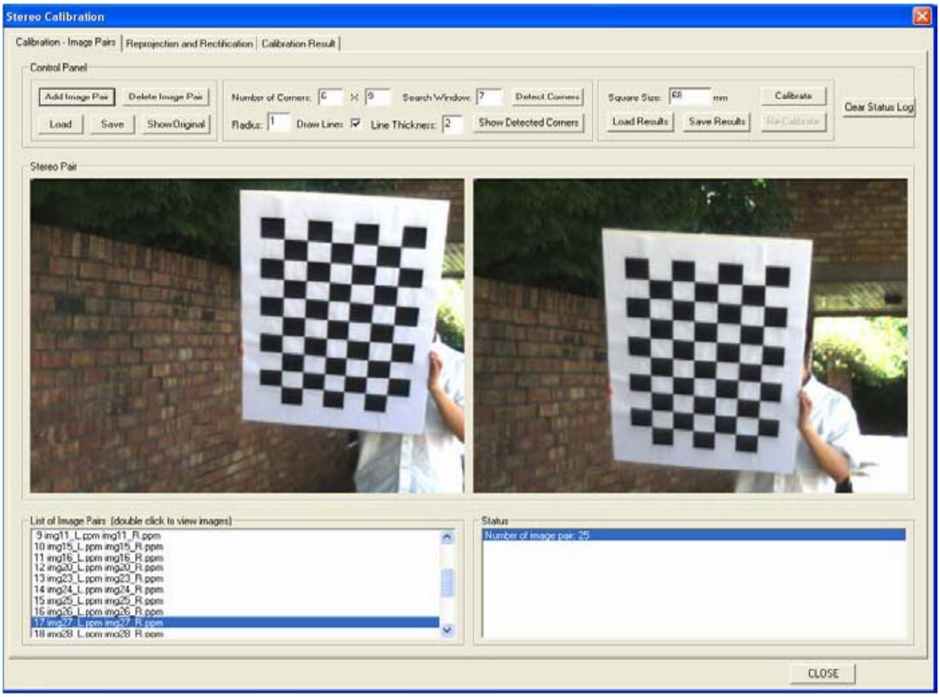


Figure 2.7 Stereo camera calibration method.

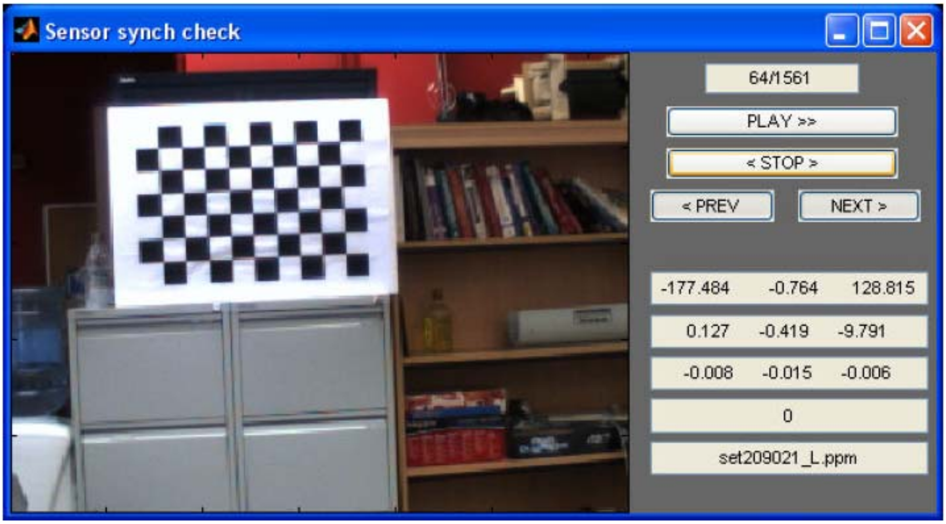


Figure 2.8 Tool for eyeballing image capture from the head position sensor

2.4.2 Image segmentation

Figure 2.7 depicts a typical image segmentation and object detection scheme. Video signals are split into individual frames using wavelet analysis. Each one of them is segmented into different regions. Statistical features are then extracted from these regions, which later are introduced into a trained classifier to recognise each individual region. Necessary grouping will be performed to group connected regions that belong to the same object. The training is carried out based on an image database with known object classes and labelled scenes. A neural network is one of the most popular classifiers for such a framework. During the image segmentation around 10 frames per second processing speed was achieved.

Two example images are shown in the first row in Figure 2.10. The second row shows the hand-labelled scene segmentation. Different colours correspond to different classes of objects. The last row presents the classification results using the current recognition system which is the starting inspiration for the CASBLIP system.

An important task of the Artificial Vision System is to recover object depth from a stereo camera system.

In order to perform image segmentation the graph cut based algorithm was implemented. The algorithm treats the image as a connected graph and takes into account global characteristics while optimising the objective function.

Multimodal features such as colour, texture, motion and estimated depth were used to segment images.

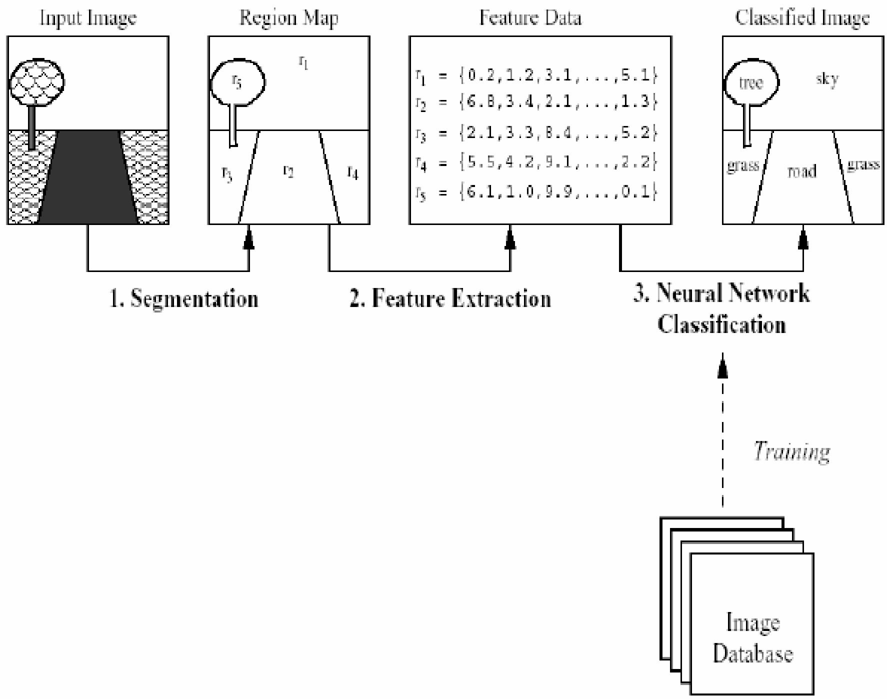


Figure 2.9 Typical image segmentation and object recognition scheme

Figure 2.10 Example of image segmentation. Each column represents the segmentation of two different images

2.4.3 Depth estimation

The most common method for estimating depth from any image is the stereo vision. Usually in the stereo vision, the depths are estimated by triangulation using two images.

With the camera calibration given for all viewpoints of the sequence, we can proceed with methods developed for calibrated structure from motion algorithms. The feature tracking algorithm already delivers a sparse surface model based on distinct feature points. This, however, is not sufficient to reconstruct geometrically correct and visually pleasing surface models. This task is accomplished by a dense disparity matching that estimates correspondences from the grey level images directly by exploiting additional geometrical constraints. A stereo vision is fundamentally limited by the baseline distance between the two cameras. Great disadvantage of the depth estimation is appearing for large distances, because even very small triangulation estimation errors translate to very large errors in distance.

Figure 2.9 illustrates the basic idea of how to estimate the object depth from known camera positions and configurations. Points A and B represent two cameras, and PA and PB are two camera planes on which scenes will be projected.

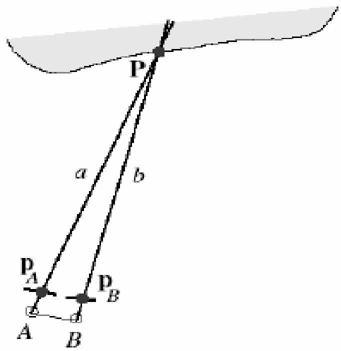


Figure 2.11 Depth estimation method

Figure 2.12 Depth map extraction from stereo images either side

In order to estimate the depths, video sequences were pre-processed and rectified. Epipolar geometry-based depth estimation is performed to produce dense depth maps, at 10 frames per second.

With a view to countering the bleeding present in the depth maps for low textures areas anisotropic diffusion was employed. Anisotropic diffusion is essentially a method of reducing noise which preserves true edges. A smoothing filter was applied for reducing noise and blur edges. The dynamic programming based estimation method has been used for the camera calibration. 10 frames per second processing speed have been achieved. Absolute Walking Time results from Mobility Test II. The eleven clusters of bars show results for the ten subjects as well as the average across subjects.

2.4.4 Object detection

Automatic detection of objects in images is a crucial and main challenge in computer vision and pattern analysis. An important application is robustly hypothesizing and verifying obstacles for safety applications in intelligent devices.

The automatic object detection has been widely studied for vehicle application and artificial vision for robotics. In our case, the automatic and real time object detection is used for artificial vision for blind people. The obstacle detection of the Real–Time Assistance Prototype, the Artificial Vision System has three major challenges:

1. The algorithm must run in real time with minimum delay in reaction time.
2. Obstacles must be detected at large distances, in order to warn the blind user about obstacles as early as possible.
3. The position and horizontal dimension of the obstacle must be estimated precisely to safely guide the blind user.
4. The algorithm must work both for static and for dynamic objects.

The first two challenges demand an algorithm able to detect obstacles in the focus of expansion where optical flow displacement vectors between consecutive frames are extremely small. The third challenge requires robust verification of obstacle boundaries.

Nowadays, obstacle detection can be realized using two technologies:

- By using stereo vision.
- By using monocular vision.

The first method, which represents the traditional vision based object detection, is realized through depth estimation from the stereo systems. This method works well and it is used in almost all technologies.

The second method using monocular vision employs a-priori knowledge and other methods and algorithms based on the relative image motion. Object detection using monocular vision is a desirable alternative.

The most common approach for detecting a class of object is to slide a window across the image and to classify each such local window as containing the target or background. This approach has been successfully used to detect rigid objects including faces, humans, cars etc. It has also been applied to articulated objects such as pedestrian. The pedestrian detection is based on the sliding window classifiers to detect object parts, and then to assemble the parts into a whole object. Another approach is to extract local interest points from the image, and then to classify each of the regions around these points, rather that looking at all possible sub windows.

In the present work, the main objects to be detected were humans, buildings, cars, motorcycles, bicycles, trees, animals, etc… and free spaces.

The system represented the detected object by a window, as shown in Figure 2.10, where the window is presented with yellow colour.

An important task of the system is the detection of a free space. This task was developed with the idea of simplifying the environment image, placed in the direction of view of the user. When a multitude of objects appear in the area of view of the system, the user will receive a sequence of sounds representing these objects. In order to simplify the user image, the idea of detecting free space was developed. In Figure 2.10, in the second image, the free path detection is represented with a light red colour. The method is based on the bounding box detection. In order to estimate the distance between the object and the system and the free path distance, the depth map method was used.

Figure 2.13 Object and free space detection

The integration of the aforementioned systems, i.e, Sensor system and Artificial Vision System, into a single unit, intends to provide more information about the surrounding environment to the blind user. Using both systems simultaneously, the user will be able to detect the obstacles in an area between 0,5 and 15m. The sensor system of the Acoustical Prototype will provide precise information on the nearest obstacles up to 5m, whereas the artificial vision will provide the information of the farther environment up to 15m. The obstacle direction of movement is indicated by acoustical sounds, in such a way that the sound moves with the same speed and direction as the obstacle.

2.5 Acoustical System

Information of the 3D environment is perceived through the eyes and ears. When the acoustic channel does not work, it is still possible to orientate using visual information (Klatzky R. L., Marston J. R., Guidice N. A., Golledge R. G., Loomis J. M. 2006). However, when only this latter is blocked, comes or auxiliary elements must be used, due to the inherent limitations of the acoustic channel.

The sensory modules provide two main types of data on the user frontal scene: one – the location, direction of the objects, and second – the set of coordinates where a horizontal plane passing at the eyes level cut the surface of the existing object. The Audio Interface is able to synthesize on real time the set of sounds to be delivered, by means of a convolution operation between every spatial filter provided by the sensor system. The Audio Interface presents audio information to the user, representing a limited area of the subject frontal scene. This area consists of a plane horizontal to the user head, located at the height of the ears. The used sounds are very short and impulsive sounds, a set of clicks processed for this system. The information is transmitted to the user via headphones.

The Acoustical System performs the next tasks:

- Perception of the depth maps sent by the Artificial Vision System.
- The calculation of the corresponding sound presentation storage of the received stereo pixels collection in a random and different manner for each image. This approximation avoids the sonar patrons which distracts the human attention from the acoustical sounds.

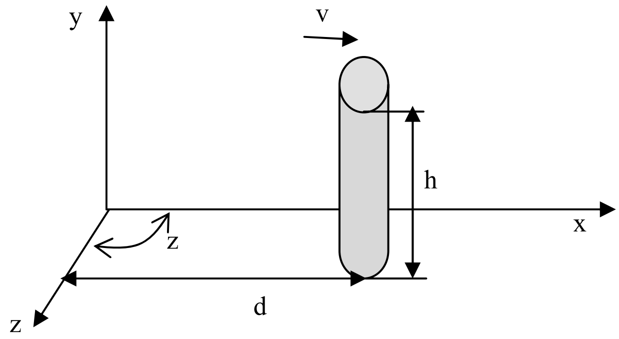


Figure 2.14 Spatial coordinates of the object

- Sound reproduction corresponding to each stereo-pixel, one by one, maintaining a predetermined separation time. In this work, three type of sounds, short without tonal qualities, are considered.

As previously mentioned, the acoustical system is in charge of the representation of virtual information such as:

Object distance (d) – the distance is inverse proportional to the sound frequency. The increased frequency of the sound represents the nearest location and the lowest sound frequency represents the farthest location.

Object direction (z) – the acoustical sounds produce a virtual run through some coordinates to other, representing a displacement in a specific travel direction where the azimuth measurements are used.

Object height (h) – the elevation coordinates sound specifies the object height.

Object speed – is proportional to the pitch changing speed.

Scene complexity – the acoustical system represents only one obstacle which is detected by the artificial vision system and classified as the nearest obstacle.

Moving object – the transmission rate is proportional to the object distance. This means that when the object gets nearest with respect to the user, the inter-click interval decreases from 100ms up to 25ms.

Visual deficit in the blind people is compensated by auditory processes (Dufour et al., 2005). The spatial sounds (Makous and Middlebrooks, 1990), (Takahashi and Keller, 1994) help humans to orientate, detect and localize the sound sources in the environment. Two major binaural cues such as Interaural Time Diference (ITD) and Interaural Level Diference (ILD) are used for sound localization. The ITD and ILD cues are the predominant cues for the localization in horizontal planes (Brungaut et al., 1999). The head rotation has also an important contribution for improving the localization.

The present section describes the Acoustical System, whose objective is the use of virtual or spatialized sounds to represent the objects location in the environment. In the present study, the spatial sounds and auditory cues were directly evaluated as a virtual navigation task. In order to achieve the proposed objective, Head-Related Transfer Functions (HRTF's) have been used. HRTFs represent the transfer characteristics of a sound path from a sound source to the listener ears (Blauert, 1995).

In order to complete this aim, it is necessary to define the following tasks:

- Research on intensive HRTFs acquisition.
- Fine tuning of the robotic system.
- HRTFs measurement system.
- Acoustical characterisation of the HRTFs measurement room.
- Improvement of the HRTFs acquisition strategy, if it is necessary.
- Research on the auditory coding of the pixelated scene.
- Preliminary perceptual testing of an early version of the M1 prototype.
- Research on the auditory coding of the distance dimension.
- Experimental design.
- Acoustical characterisation of the experimental room.
- Research on the M1 device auditory coding strategy.

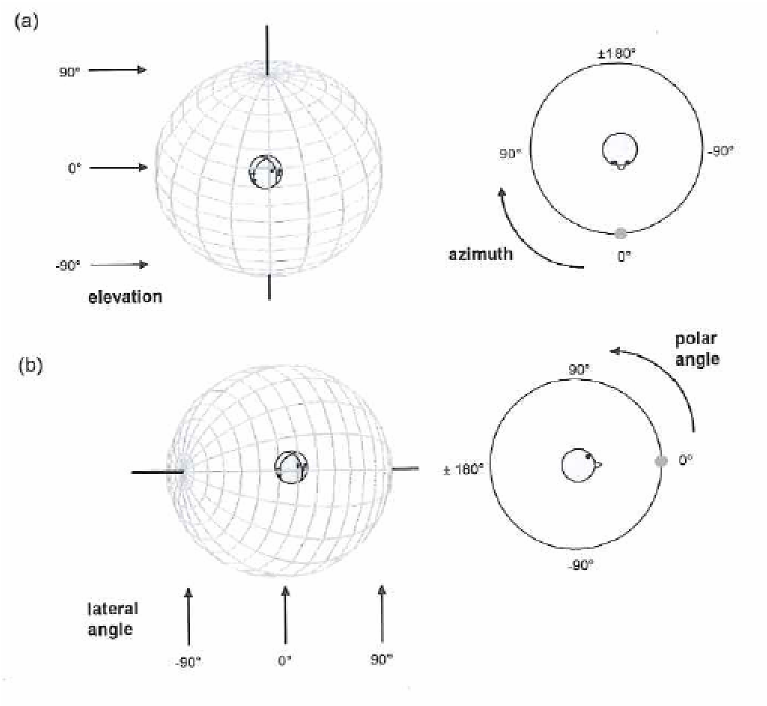


Figure 2.15 Spatial coordinate system a) represents the azimuth/elevation coordinate system consisting of a single pole oriented vertically. Azimuth describes the horizontal angular displacement due to rotation around the pole, with 0 º azimuths at the front and 180º and -180º azimuths at the ear. b) describes the lateral/polar coordinate system. The system consists of a single pole passing through the two years.

2.5.1 Stimuli

The effect of the environment, including the head, ear, etc, is assumed to be linear, locally time-invariant transformation of the signal. So that the received signals at the ears are linearly related to the source stimuli. This transformation from source to ear for a given direction and environment is conveniently described by a transfer function in the frequency domain. For a better perception of the spatial sounds in 3D space, it is required to process the sound with the Head Related Transfer Function (HRTF). A single specific Head Related Transfer Function is defined to be a specific individual's left or right ear far-field frequency response, because it is the measurement of a specific point in the free field to a specific point in the ear canal. Usually, the HRTFs are defined empirically as the Fourier transform of the head-related impulse response HRIR. For reverberant environment, the term HRTF is reserved. Usually, the HRTFs are measured in a KEMAR mannequin for both left and right ears, or in human ears. As mentioned before, the HRTFs are measured for different spatial point's azimuth for left and right directions, elevation for up and down directions and finally the distance for frontal, lateral and back planes.

Three different sounds were used for the development of the system. The delta sound, a Synthetic Percussive sound and a Modified Synthetic Musical sound (Marimba). At the left side the time domain is represented for the left and right band at the right side the sound spectrum is presented.

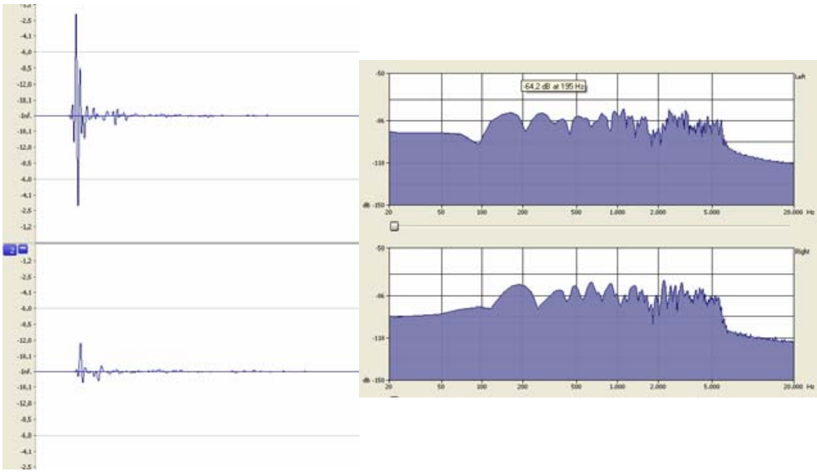
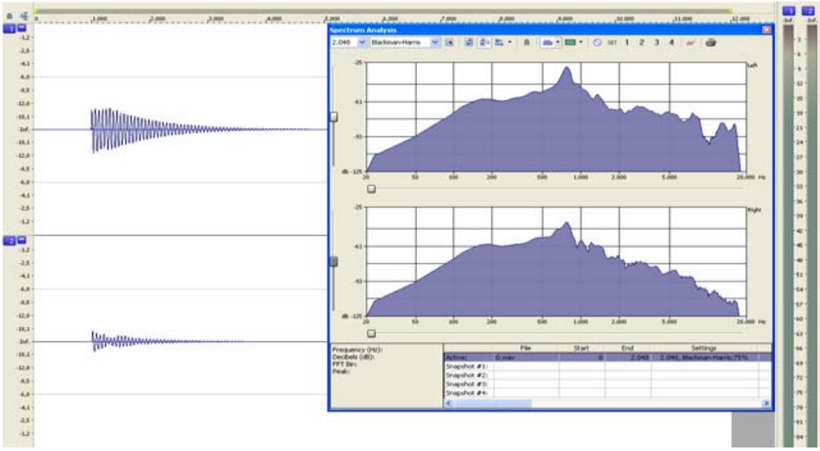


Figure 2.16 Signal spectrum for the Delta sound

Figure 2.17 Signal spectrum for the Synthetic Percussive sound



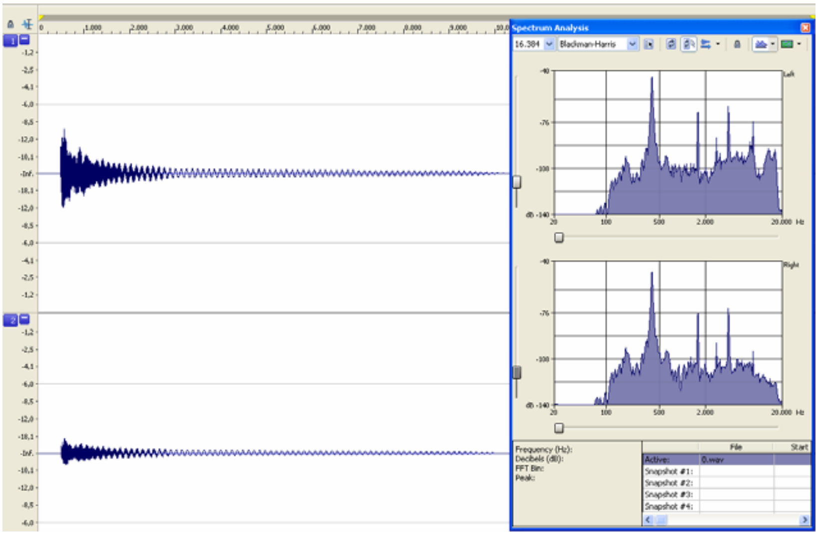


Figure 2.18 Signal spectrum for the Modified Synthetic Musical sound (Marimba)

2.5.2 HRTF's measurement

In order to measure the HRTFs for the Acoustical System, the technique for measuring individual HRTFs for each individual of the system was developed. In order to measure the HRTFs, a robotic arm moved the speakerphone at the corresponding position, in accordance with the user head. A delta sound was then reproduced through the speakerphone (this sound is named MLBS (Maximum Length Binary Sequence)). The response of the microphones which are situated in the manikin or user ears is registered in a data base. The process was repeated for each spatial coordinate necessary for the experiments. The HRTFs were measured for each needed coordinate, since it was required to obtain as much possible real spatial sounds. Note that the sounds for these tests were not generated by the computer as in the precedent chapter. In this case, the sounds were used for a real navigation system for blind people.

Recording of HRTF's was carried out in an anechoic chamber (Figure 2.12) of $25m^2$. The robotic system is able to move the loudspeakers at any location in a radius of 5 meters.

Due to that the HRTF must be measured from the two ears, there is necessary to define the two inputs and output signals. Lets $x_1(n)$ be the digital register of the sound that must be reproduced by the speakerphone. Lets $y_1(n)$ be the final register recorded by the microphone placed in one of the acoustic channels of the manikin or man, corresponding to the response to $x_1(n)$. Similarly, let $x_2(n)$ be the sound to be reproduced through the headphone and $y_2(n)$ the answer registered by the headphone, respectively for the second ear. The location of the head in the room is assumed to be fixed and is not explicitly included in our explication.

In order to determine $x_1(n)$, it is necessary to generate a $x_2(n)$ such that the $y_2(n)$ is identical to $y_1(n)$. In that way, we achieve that an acoustic stimulus generated from the speakerphone and another generated by the headphones, produce the same results in the auditive channel of the user or manikin. Therefore we obtain the same acoustical and spatial impression. In order to obtain these stimuli, a digital filter which transforms the $x_1(n)$ into $x_2(n)$ has been developed.

In the transformed frequency domain, let be X_1 the representation of the $x_1(n)$ and Y_1 the representation of the $y_2(n)$.

Figure 2.19 The anechoic chamber with small reverberation level and the robotic system. The walls and floor of the chamber are lined with sound absorbing wedges. The whole robotic system can be moved to place the loudspeaker at any location in radius of 5 meters. A chair for manikin is represented in a fixed platform.

Then Y_1, which is the registered response of the $x_1(n)$ reproduction, is:

$$Y_1 = X_1 LFM \tag{1}$$

In (1), L represents the grouped transfer function of the speakerphone and all audio reproduction system. F represents the transfer function of the environment situated between the speakerphone and the additive channel (HRTF) and M represents the set of functions composed by the microphone and the whole audio reproduction system.

The response registered by the microphone via headphones, when the $x_2(n)$ is reproduced, can be expressed as follows:

$$Y_2 = X_2 HM \tag{2}$$

where H represents the transfer function of the headphone and all reproduction system to the additive channel.

.

If $Y_1 = Y_2$, isolating X_2 we obtain:

$$X_2 = \frac{X_1 LF}{H} \tag{3}$$

Then, for any measurement the digital filter will be defined as follows:

$$T = \frac{LF}{H} \tag{4}$$

Therefore, it will filter the signal $x_1(n)$ and the resulting signal $x_2(n)$ will be reproduced by the headphone; then the signal registered by the microphone, which is placed in the auditive channel must be $y_1(n)$. This signal must be equal to the signal $x_1(n)$, which is reproduced by the speakerphone.

The filter described by (4) describes the speakerphone for a single spatial position for only one ear. For both ears two filters are required for the simulation of each signal source for a determined spatial position.

Assuming that we measure the Y_1 and X_1 transfer functions for different spatial positions for both ears at the same time, the Transfer Function speakerphone-microphone (G_{LM}) is defined as follows:

$$G_{LM} = \frac{Y_1}{X_1} = L \cdot F \cdot M \tag{5}$$

Having the function given by (5) simultaneously for both ears, we measure both transfer functions Y_2 and X_2, on which the transfer functions headphone-microphone G_{HM}, are defined:

$$G_{HM} = \frac{Y_2}{X_2} = H \cdot M \tag{6}$$

The necessary filters for the sound simulation are obtained from the function speakerphone-microphone G_{LM} for each ear, as the reverse of the function headphone-microphone G_{HM} of the same ear (see (4)). So, for both ears:

$$T = \frac{G_{LM}}{G_{HM}} = \frac{L \cdot F \cdot M}{H \cdot M} = \frac{L \cdot F}{H} \tag{7}$$

For both transfer function speakerphone-microphone G_{LM} and headphone-microphone G_{HM}, the measurement technique of the impulse response Maximum Length Binary Responses MLBS was applied with later crossed correlation between the system answer and input of the MLBS.

The impulse response of the system can be obtained through circular crossed correlation between input MLBS of the system and the output answer. This is, if we apply to the system an MLBS, which will called *s(n)*, and measure the output the signal *y(n)* during the time which MLBS lasts, the impulse response *h(n)* will be defined as follows:

$$h(n) = \Omega_{sy}(n) = s(n)\Phi y(n) = \frac{1}{L+1}\sum_{k=0}^{L-1} s(k)\cdot y(n+k) \qquad (8)$$

where Φ represents the circular or periodic crossed correlation operation, corrupted by the aliasing time, and not a pure impulse response.

In the event that the sequence is enough long, then the resultant aliasing can be rejected. Due to that, the direct implementation of (8) for long sound sequences require high computational time, the equivalent between the correlation and periodic crossed correlation has been used. The obtained information was passed into the frequency domain, where the convolution operation is translated into a vector multiplication.

After this, the results were passed into the frequency domain, where the convolution operation is translated into a vector multiplication.

$$a(n)\Phi b(n) = \frac{1}{L+1} a(-n) * b(n) \qquad (9)$$

where the inversion of the first sequence is circular, similar to the convolution. Nevertheless, the computational time results to be enough high, due to that the used Fast Fourier Transform (FFT) have a length of 2^k-1. In order to obtain an increasing performance in time processing the FFT length has to be $(2^k-1)^2$.

Finally, using the Fast Hadamard Transform (FHF), it was possible to reduce the computational time between the two magnitudes. The *h(n)* is then calculated as follows:

$$h(n) = \frac{1}{(L+1)s[0]} P_2 \langle S_2 \{H_{L+1}[S_1(P_1 y(n))]\}\rangle \qquad (10)$$

In this case to the system has been applied a MLBS s(n) with a length L, after what the result y(n) was registered. The matrix *P* is the permutation matrix, the matrix *S* is matrix of rescaling, the H_{L+1} is the matrix Hadamard of degree $L+1$.

2.5.3 Validation of the HRTFs

After the HRTFs were measured, with the equipment shown in figure 3.13, it was verified if the HRTFs are realistic and externalized. For this purpose, an off-line localization procedure was carried out.

The tests for sound localization were completed using a table of sounds interface, where each sound was located under 976 cells. Each cell represented one spatial sound at specific coordinates. Figure 2.14 shows, with orange colour, column 31. It contains the cells so that when clinking on them the sounds corresponding to the central plane will sound. From column 31 in both directions, the convolved sounds are located. On the left side, the sounds representing left side are located, whereas on the right side the sounds which represent the convolved sounds for the right plane are placed. There are 30 sounds at each side, which represent the 30 pixels of the laser and camera. For the experiment, sixteen distances were used, between 0m from the user and 5m, proportionally distributed, with a step of 0,8m. It means that the first cell represents the sound at the coordinates: 30º in azimuth and distance of 0m at the eyes level. Since the sensors are placed in a horizontal plane, the acoustic test used the sounds for the horizontal plane. Imaginary, a table where on each cell is specific sounds which represent the horizontal plane of the sensory system.

At the top part of the interface, two windows are placed. On the left side, the window showing the mouse position is placed, whereas on the right side, the window where the user can select the sound file is located. The buttons at the bottom of the interface enable the Resolution Test, Archive Load, Play, Stop and the mode of sound sounding, i.e, using sequential sound or a mouse click.

During the tests of HRTF's and stimulus validation the stimuli were presented through the earphones.

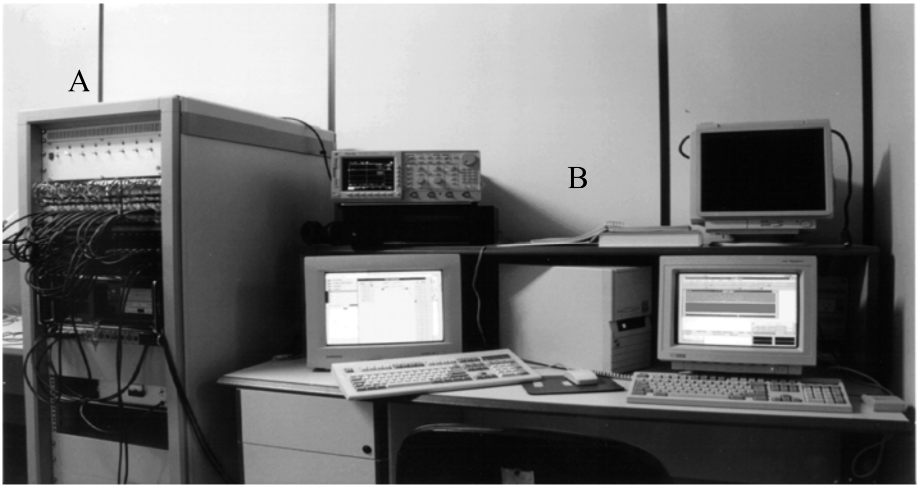


Figure 2.20 Sound generation and processing system. A represents the Huron system based on 40 analogical outputs and 8 analogical inputs and 8 DSPs 56002. The system was used for HRTFs measurements and experiments for mono and multi sources localization. B represents the computer where the HRTF's were processed and off-line processing of the spatialized sounds performed.

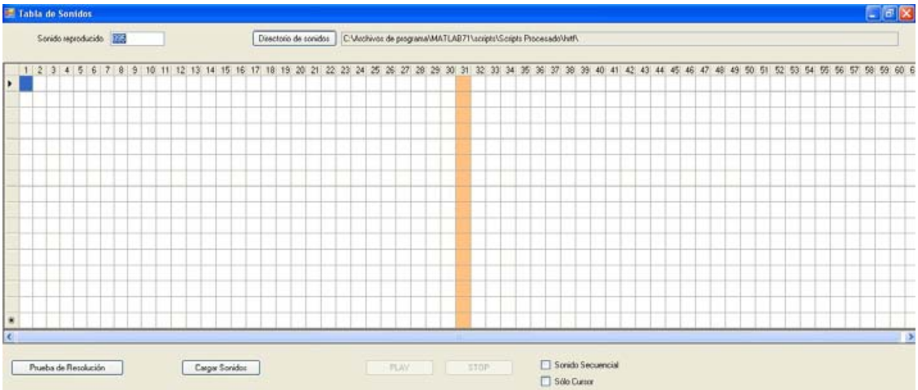


Figure 2.21 Test interface

2.5.4 *Sonification strategy*

The general approach for the sonification strategy consists of sonifying the real environment using trains of spatial short sounds. These sounds are known to be very locatable, to represent the different types of information provided by the video system. Different types of information will be delivered simultaneously frame by frame by the vision system. Trains of sounds generated and processed with different timbres, pitches, and inter-click intervals (ICI) are used for different objects representation allowing the user to recognize them. Three types of strategies are developed: sonification strategy for moving objects, sonification strategy for free path detection and sonification strategy for two-dimensional vision (sensory).

Sensor sonification strategy

The main aim of the sensor sonification strategy is to create the sound module which will be used for the object localization perceived by the sensor system. A map of a train of short impulses was created for a two-dimensional frontal plane with 32º in azimuth at the left and right side from the user direction of view and for a distance between 0,3m and 5m. The sounds were generated every degree in azimuth and every cm in distance. Each sound is saved into an archive, in which each name represents the polar coordinates; x represents the azimuth and the z represents the distance.

This practically means that if the sensor system detects an object of 20cm in thickness at distance of 1m in the centre of view, then 20 sounds at the corresponding position, 10sounds from the centre-left and 10 sounds from the centre-right at the distance of 1m will be delivered by the acoustical module and reproduced by the system.

The strategy is to sonify every pixel of the image; in this way, the user will obtain a whole image of the detected object. .

Moving object sonification strategy

A train of very short sounds represents the moving object perceived by the stereo-vision system. The strategy is similar to the one used for representing the sensor information previously described. In order to force a difference between the sounds used for the sensor system and the stereo-vision system, the moving object sounds were generated with a different timbre. Every frame is sonified including the object 3D spatial position**,** the object direction relative to the user and the object speed relative to the user.

The spatial position of the object is given by the coordinates x (azimuth) and y (elevation) corresponding to the center of their bounding boxes. The distance, or z coordinate, is represented by the closer coordinate provided by the system. In addition, the pitch of the sound changes progressively with the distance. According to the results of the survey on blind people, spatial coherent auditory information on the moving object is provided for a distance range between 0.5 and 15 m. Beyond 15 m, the object is represented by a spatial sound which corresponds to the last distance of the cited range (this is, 15 m). This is expected to allow the user detecting the presence of the moving object, when it is detected by the sensor subsystem also for a further distance.

The object direction relative to the user, represents whether the object moves or not towards the user. Lowest ICIs are used when the object is facing completely the user, and the highest ICIs are used when the object has completely divergent trajectory. Therefore, the ICI is expected to be a main feature to indicate the user whether there is or not a risk of collision.

The object speed is codified by the speed of changing the pitch when passing from one distance to the adjacent one.

Free path strategy

In order to facilitate the sonification strategy for the free path, a very simple method was selected. In this case, only the borders of the path are sonified. For the free path, a new folder train of sounds with a particular timbre was created which differs from the previous timbres and that can be recognized easily by the user as “border of a free path”. Due to that the free path detection is obtained by the stereo-vision system; the maximum distance is detected at 15m. This allows the user to perceive in a range of 0-15m and to perceive the borders of the free path if they exist.

Methods

– A number of blind persons will carry out the test without the help of any residual vision they may have.

– The subject has previously undergone the Basic Learning Protocol.

– The test will be carried out in three stages: Pre-training test, training sessions and post-training test. They will be appropriately described after the detailed description of the test.

– Which parameters can be measured?

– The time needed to carry out the task.

– The number of hits.

– The trajectory followed by the person, this is, whether it is the correct trajectory or a larger one: steps backwards, steps by the side of the set up instead of between each pair of objects, etc. This makes sense only if we allow the person to move freely without correcting her about any wrong followed direction. Regarding the question of how to register it, may be with a camera located at a high point?

– The ability to localize every object. We already have an indirect measurement from both the number of hits and the speed. Nevertheless, a direct measurement could be done before beginning the task, for example when carrying out the Learning Protocol. Any deviation from the expected pointing direction should be computed. It can be done in a gross qualitative way: "non deviated pointing", "deviated to the left", and "deviated to the right". Could be used a camera located at a high point?

Detailed description of the test:

A series of pillar shaped obstacles are located in a large enough space, where no any other object exists (See Figure 2.22).

- They are laid out as pairs of objects, one to the right and one to the left.

- The space between every pair of objects defines the free path to be stepped through.

- The objects are, at the same time, both obstacles to be avoided and references to find the free path.

- Every pair of objects defining the path should be focused in such a way that they can be simultaneously perceived inside the field of view of the prototype.

- No any detectable object or surface must be in the proximities of the experimental setup.

- Dimensions, shape and location of the objects.

- The environmental sound should be as low as possible, and the reverberation quality of the space should be also similar in any place where the test was carried out.

The subject must walk throughout the free path trying not to hit the objects.

After being told which the task to be carried out is, the person must be located at the starting point, facing a silent area.

Then she will be told “start”, and the time must be computed since that moment.

She must turn the head until detecting one object. Se must try to perceive where it is (exercise 1 of the Basic Learning Protocol). Next, she must try to detect another object at another location.

The subject must be told that the best way to perceive the free path is trying to perceive both objects at the same time, this is, everyone located at each side of the subject (exercise 2 of the Basic Learning Protocol).

Once the subject has detected where the two closer objects are, she must walk right towards the free space between the objects, trying to over pass them without hitting them.

Once the subject has overstepped both objects, the person should stop for a while and look again for a new pair of objects which are defining the next free path to walk throughout it.

The person must maintain in memory the location of the overstepped objects in order to avoid using them as the new pair of references. This way, we prevent the person to follow a wrong way using one of the overstepped objects a reference.

It could happen that more than 2 objects appear within the field of view. The path is defined by the two closer objects.

It may happen that the person goes backwards. If we do not want to prolong the test, the person should be told that “you are going in the opposite direction”. Another possibility is allowing the person to move between the objects until she finds the end.

If the person faces a silent area, then it means that she is using as reference an overstepped pillar, and then she must go on looking for the correct pair of reference objects.

The exercises finish when the person arrives to a place where an “end-object” is detected. She must touch it, and then timing must finish.

The end-object may be a single surface. If so, we are including in the task a new perceptual task, which is distinguishing between single columns or pillars and a wider surface, this is, the perception of the width of the object (exercises 3 and 4 of the Basic Learning Protocol).

Pre-training test

The subject must be told which the task to be carried out is (after discussing the experimental test proposed, we should elaborate a letter with all the instructions to be read to the blind participant).

Then, she must proceed according to the instructions (it is very important to remind the person to keep the head looking at the front, not to the floor)

Training period

– The training period will start immediately after the pre-training session.

– Number of training sessions: 3 in different days

– Estimated time for every training session. This time should be the same for all the participant subjects.

– The aim is to offer the person a guided interaction with the environment, in the sense of, for example, asking her to perceive every object independently, both listening and by touch, and then to perceive both objects at the same time, etc.

– Every training session may consist of:

– The person receives a verbal feedback, this is, a verbal description of the set-up with the location of the objects (4 pairs of objects; given a pair, each one at each side, right and left).

– The person is reminded to focus the two closest objects at the same time.

– The person begins the walk and we invite the person to approximate to the objects and to touch them.

– We will remind the person that the closer objects sounds closer than the further ones.

– The person may perceive more than two objects. The person must take as spatial references the two closer objects.

– The walk is repeated twice.

Post-training test

After training, if we maintain the same set-up the blind participant can use many other cues than the auditory ones to carry out the task. Given that what we want to measure the performance when using the prototype, another set up should be presented. In order to keep the difficulty level of the task to be carried out as similar as possible, a left-right symmetric set up is proposed (see Figure 2.22).

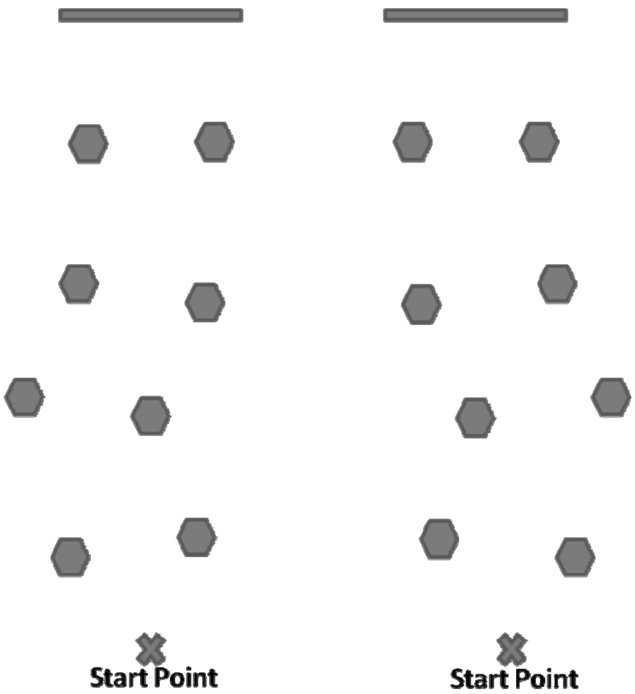


Figure 2.22 Test trajectories

2.6 Communication Interface

The communication interface is based on cross-platform software for image and digital signal processing and synthesis. It is implemented as a C++ class library. Each one of the sub interfaces are written in a programming language C++ separately and converted to a DLL. The communication interface is designed to work for Windows XP operational system in conjunction with the FPGA device. The program sets up the graphs, data window and video image window.

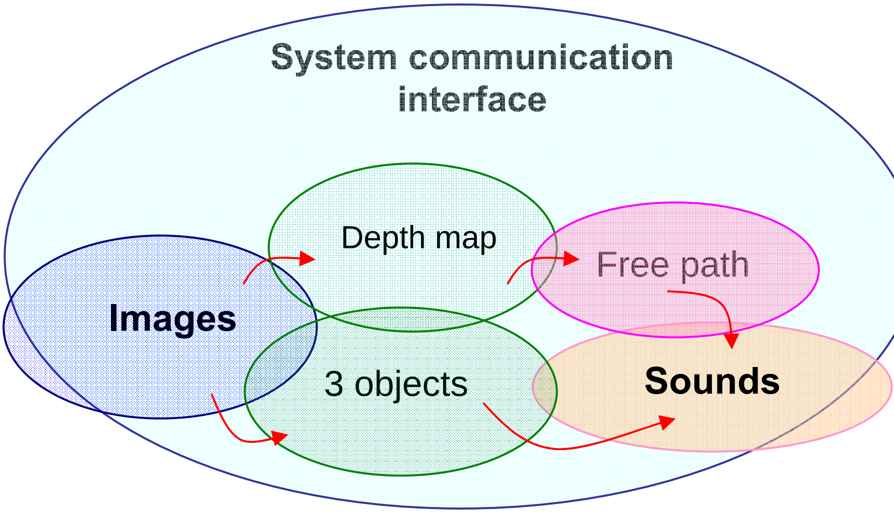


Figure 2.23 System integration components

The source code of the communication interface is organized in the following interfaces:

2.6.1 Image processing interface

The image processing interface aim is to receive through two cameras the environment image, to process the frames (right and left) and extract the depth maps, from which the 3 detected objects and their features are listed.

The objects field contains a list with the 3 segmented objects, providing their average depth (object distance), their speed, the bounding box they are contained in, an integer which ranks the objects between 1 and 3 in terms of their hazard, an object identifier (for tracking) and a frame specification.

The software recognizes the depth map as a field of width × height array of linear distance values. A value of 0 is reserved for "no distance measured/available". The depth map is specified from left to right and from top to bottom.

CASBliP Object structure definition:

The image processing interface will communicate with the system communication interface, providing the depth map and the information about the 3 detected objects as follows:

```
int width, heigth; // Depth map image size
double near, far; // Depth map distance measurement range, in meters
double timeStamp; // time stamp measured in seconds
int *depthMap; // Depth map vector
int numObjects; // number of detected objects (up to 3)
int averageDepth;// Average depth of one object
double speed[3]; // x, y, z of speed vector
int x1, y1, x2, y2; // Bounding box
int hazardness; // Hazard level
int objectID; // Id of the tracked object
```

For a NULL **pointer** as no depth maps or objects are returned for the first few frames of the DLL running.

```
while(1){
	int d = lpfnDllFuncNextFrame((void**)&casObj,
(void**)&imgD,(void**)&imgS);
	/*printf("Depth = %d\nBox Coords (x1,y1)->(x2,y2) = (%d,%d)-
>(%d,%d)\n",
```

```
                              casObj->averageDepth, casObj->x1, casObj->y1,
casObj->x2, casObj->y2);*/
              if(casObj != NULL){
              printf("\n\nObj Coords (X1, Y1, depth) = (%d,%d, %d)\n, width =
%d, height = %d, near = %lf, far = %lf\n\n speed = %lf,%lf,%lf\n",
casObj->XObj, casObj->YObj, casObj->depthObj,
casObj->width, casObj->height, casObj->nearDepth, casObj->farDepth,
casObj->speed[0], casObj->speed[1], casObj->speed[2]);
}
                              // Save Images to look at
                              //sprintf(fname,"depthMap%s%d.pgm", (cnt <
10)?"0000":(cnt<100)?"000":(cnt<1000)?"00":"0", cnt);
                              //printf("Saving %s and ", fname);
                              //cvSaveImage(fname, imgD); // NOTE THIS WILL
LOOK BLACK (adjust contrast / brightness in image viewer to look at it)
                              //sprintf(fname,"segmMap%s%d.ppm", (cnt <
10)?"0000":(cnt<100)?"000":(cnt<1000)?"00":"0", cnt);
                              //printf("%s\n", fname);
                //cvSaveImage(fname, imgS);
                              // TO get pixels imgD->imageData

                              cnt ++;
                         }
```

2.6.2 Sound interface

The sound interface represents a software platform created in program language C++, which contains a C++ library. The interface is created for sound delivery for a specific coordinate system. Each sound from the library represents a specific spatial sound for a specific coordinate and for a specific object. For this purpose, a library for three different objects was developed with three types of sounds. The sound interface is created in a DLL format which is defined in:

SoundModule.lib
SoundModule.dll
SoundModule.h

Image processing module will communicate with ULL module using SoundModule.dll, which uses the following structures:

Initialization & Finalization

The sound module will be initialized using the initialize() function. An OperatingParameters struct is passed indicating the distance measurement range in meters, since depth maps are not going to be integrated in the Real-Time prototype. The function will return an error value or NO_ERROR.

The sound module is ended using the finalize() function.

```
namespace SoundModule {
	enum {NO_ERROR=0, UNSPECIFIED_ERROR};
	struct OperatingParameters {
		double near, far; // Range in meters
		bool debug;
	};
	int initialize(OperatingParameters&);
	void finalize();
}
```

Communications

For each processed frame, the Image module will call the newFrame() function, from the sound module, which is the main function of the sound module and will trigger the strategy and all the sonification stages. The call to newFrame() is blocked for the frame processing but not for the sound reproduction that will be sent to the headphones in background. The function will return an error code or NO_ERROR.

The function will receive a time stamp measured in seconds and a frame specification. The Frame structure contains this frame specification. It contains a list of 3 objects and a free path.

The objects field contains the list of the 3 segmented objects and the free path, providing their average depth (object distance), their speed, the bounding box they are contained in, an integer which ranks objects 1 to 3 in terms of hazard and an object identifier (for tracking).

Each object can be defined in the type field as UNCLASSIFIED for the object segmentation, and PATH classification work. In case of having a free path, the bounding box integers will only contain x1 and x2 values (y1 and y2 will be 0, since free paths are indicated in horizontal axis).

```
namespace SoundModule {
	typedef unsigned char DepthValue;
	enum ObjectType {UNCLASSIFIED, PATH};
	struct Object {
```

```
        int averageDepth;
        double speed[3]; // x, y, z of speed vector
        int x1, y1, x2, y2; // Bounding box
        int hazardness; // Hazardness level
        int objectID; // Id of the tracked object
        ObjectType type;
    };
    struct Frame {
        int numObjects;
        Object *objects;
    };
    int newFrame(double timeStamp, Frame&);
}
```

2.6.3 Communication interface design

The communication interface has been designed taking into account the user limitations and necessities, taking into consideration that the users are the blind and partial sighted people. The colour and graphical design was taken into account. The main characteristics of the interface are:

- Simplicity. The interface should be represented in a simplest form, to avoid the excess of information as text, buttons, publicity or copyright.
- Functionality. The interface must contain the main buttons. That is why the interface contains only the buttons of START and STOP of the program and EXIT of the whole system.
- Colour– the colour of the interface was selected according to their characteristics.

The grey colour represents the seriousness and simplicity.

The blue colour generates the tranquillity, relax, freshness. Besides, it is a colour which transmits authority, respect, dignity and decency. Its clear tones refer to confidence, vivacity and power.

The blue colour represents the innocence, purity, honesty, peace and tranquillity.

The orange colour is an open colour, a receptive and informal colour which represents the concepts as glory, progress and vanity.

Yellow colour produces pleasant sensations; transmit tranquillity, luminosity, vitality and progress.

Brown colour represents the honesty.

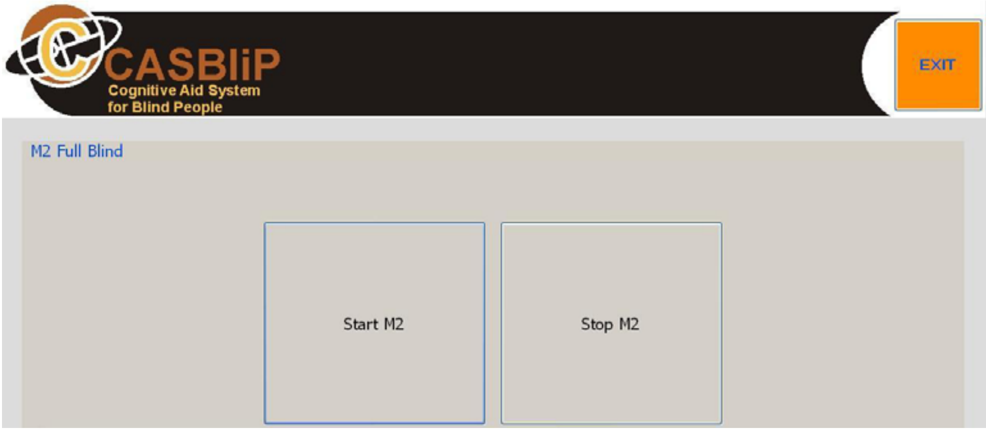


Figure 2.24 Communication interface design

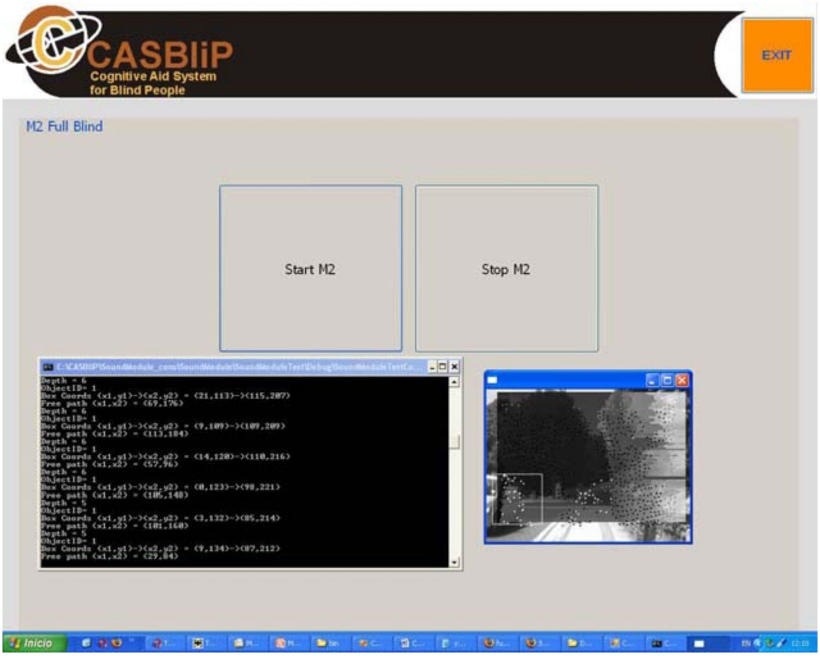


Figure 2.25 Communication interface design

The combination of open colours with the colours as grey, black, and white represent the high technology.

- Typography. An important element of the graphical design of the system is the used typography. Similarly to the colour, the typography represents special particularities.

When using the electronic format, it is important that the typography is not blurred; it must be clear. Also the typography body plays an important role in the design.

The whole graphical display is based on two sections see Figure 2.17:

1. the head of area
2. the working area

The head area consists of the system logo and anagram and the EXIT button. Pressing the EXIT button the system switch off the software.

The working area is based on one window placed in a horizontal form. The window is named "***M2 Full Blind***". It contains the standard START and STOP buttons. Pressing the START button the system opens a new interface where, beside the START and STOP buttons, two processing windows appear (see Figure 2.25).

3. CHAPTER III: SPATIAL SOUNDS

CHAPTER III: SPATIAL SOUNDS

3.0.Summary

The present chapter describes the auditory spatial perception and the human auditory system. This chapter is divided in two sections. In the first section, an overview of the auditory system is described including its basic components and auditory organisation. This section provides a background for the second section of the chapter, which discusses sound localization with acoustic cues (monaural and binaural cues, interaural time difference and interaural level difference, reverberation effect, cone of confusion, precedence effect and cross-correlation model) for sound localization. This review is aimed to represent the whole picture of the level of accuracy for sound localization. This chapter introduces the foundations of the next chapter, in which the temporal properties and efficiency of the sound source localization are analyzed. Furthermore, within the context of this thesis, the discussion of auditory perception has been carried out with the purpose of telling what is important for human listeners in the spatial field and what needs to be reproduced. As it will be seen in these chapter, the dominant cues for sound localization are the differences between the two ear input signal.

3.1 Human ear

Human ear is a complex system which involves physiology, psychology and acoustics. The human ear developed as a survival system, caring on the recognition, localization and communication of the detected sounds both noises and speech. It serves as a transducer which converts sound waves into a nerve impulses transmitted to the brain. The ears perform frequency analysis of the sound energy allowing us to perceive the sound pitch of the wave's frequencies, the sound loudness by the detection of the wave amplitude and the sound timbre by the detection of various frequencies. The human ear consists of three main parts: the outer ear, the middle ear and the inner ear. Each of these parts serves a specific task in the sound detection and interpretation. An illustration of the human ear is shown in Figure 3.1.

The *outer ear* consists of the *pinna* (a cartilaginous flap) surrounding a resonant cavity which leads to the eardrum via the ear canal. The outer ear acts as an acoustic filter and transforms incoming soundwaves, contributing to the frequency spectrum of the pressure wave that ultimately vibrates the eardrum. This energy is transferred mechanically by the middle ear to the cochlea.

The *middle ear* is an air-filled cavity that contains three small bones called the ossicles (malleus, incus and stapes) connected in a chain. The main function of the middle ear is to perform the impedance matching. It also implements an efficient transmission of the sound energy from the air to the fluids. The combined effect of the ear canal and the middle ear shift the resonance peak of the ear from 3 kHz of the ear canal alone to 4 kHz (Hudde, 2005).

The *inner ear* is the part where mechanical vibrations are transformed to neural signals, which include the vestibular organ and cochlea. The *cochlea*, the only organ from the inner ear which has auditory function, is essentially a coiled tube, located in the temporal bone of the skull, which is divided along its length by membranes into three fluid-filled compartments. The vibration of the eardrum causes pressure waves to travel through the fluid of the cochlea, setting up travelling waves in the lower basilar membrane, which is approximately 35 mm long.

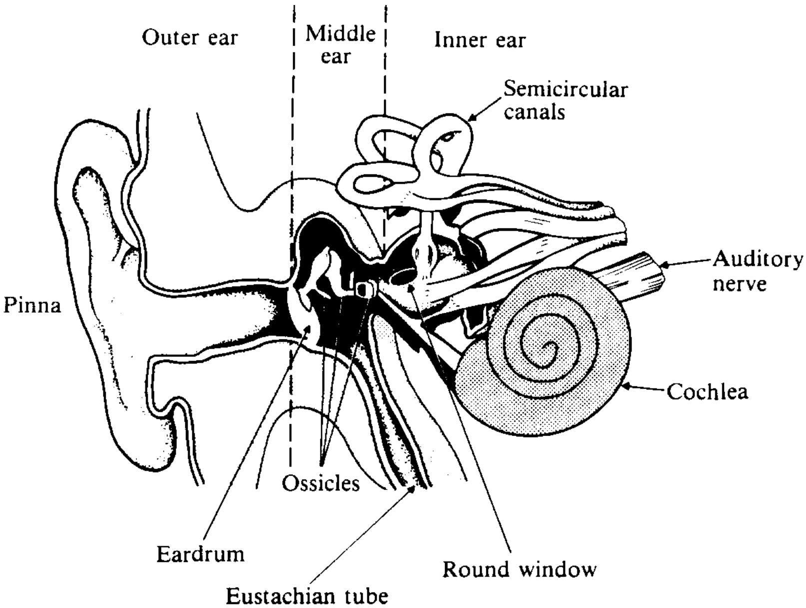


Figure 3.1 Human ear

It is along this membrane that the sensory receptors of the auditory system reside. The outer and inner hair cells are arranged in rows along the length of the membrane. The inner hair cells are the primary mechano-electric transducers of the system, converting the motion of the basilar membrane into neural signals. This mechanism involves the deflection of stereocilia located on the apical ends of the hair cells: this movement is thought to influence mechano-sensitive ion channels in the stereocilia membrane, causing voltage fluctuations within the cell. Depolarisation leads to neurotransmitter release from the basal ends of the hair cells, which contact fibres of the auditory nerve (part of the vestibule cochlear or 8th cranial nerve). The auditory nerve serves as the primary transmission line to (and from) the brain. The outer hair cells also transduce stereocilia displacement, but their primary action is to generate positive feedback forces that enhance the motion of the basilar membrane.

3.2 Sound localization

The localization of acoustic signals in space is an important function of the human auditory system. Sound source localization is affected by the head, torso and pinnae cues. The localization cues can be divided into two categories: binaural and monaural cues. The first class of cue is the *binaural* cues that arise as a result of the separation of the two ears in space. Two main binaural cues define the location of a sound source relative to a listener, and these are the interaural differences in time and level (ITD and ITD). Further to these, *monaural* cues are available from the effect of the auditory periphery on incoming sound waves (e.g. Blauert, 1983; Middlebrooks and Green, 1991; Carlile, 1996b; Popper and Fay, 1992).

Much of our insight into the cues to sound source location has come from neurophysiological and psychoacoustic studies. In these studies, acoustic stimuli are presented in a range of ways. In the most realistic stimulation paradigms, sounds are presented via loudspeakers in external space (referred to here as ‘free-field’ listening). However for a more controlled environment, many researchers have opted for headphone presentation. In this case, the percept is of a sound that is located inside the head, but binaural cues can be varied to give the impression of lateral displacement within the head, known as lateralisation. More recently, the development of Virtual Auditory Space (VAS) technology has greatly expanded the scope of research in this area. Using this approach, realistic spatial cues can be incorporated into headphone listening to give an accurate simulation of the free-field listening experience.

3.3 Monaural cues

Although humans have two ears, a great deal of information about sound location can be obtained by listening through just one ear (Plack J. Christopher book the sense of hearing, Angell and Fite, 1901). Since the incoming sound waves are modified by the upper body, head and pinnae, the monaural (spectral) cues arise on sound location. Sound localization based on monaural cues can be studied only with headphones, assuming perfect symmetry of the head. The evaluation of monaural cues depends on the ability of a listener to distinguish between the spectrum of the source signal and the effect of source position on the spectrum. In the Angell and Fite study, long trainings were required with an error of 18º. Later, Butler (1990) in his studies demonstrated that monaural listeners could not localise in the horizontal plane, but that elevation estimations were near-normal on the side of the functioning ear.

Wightman and Kistler (1997) made several experiments with monaural localization and found that monaural results were reasonable in the free-field; the localization ability was completely abolished in the virtual conditions.

3.4 Binaural cues

According to the classic "duplex theory of sound localization" by Lord Rayleigh of binaural localization, the localization is facilitated by the integration of ITDs (Interaural Time Diferences) which are used for localization at low frequencies (as it was proposed by Lord Rayleigh) and ILDs (Interaural Level Diferences) at high frequencies (in humans above 2-3 kHz) where the waveform is short and the head can act as an effective acoustic shadow. Binaural cues represent the differences between the ear input signals, when the monaural cues depend on the ability of the listener to distinguish between the spectrum of the source signal and the effect of the source position on the spectrum. Thanks to their properties, the binaural cues are more robust and more important on sound source localization. Under normal environment conditions, humans can derive location information from binaural cues with considerable precision.

The binaural cues of ITD and ILD are the most important cues for horizontal or azimuthal localization. In these previous years, a high awareness of the importance of another cue for sound localization in vertical dimension, called spectral cues has arisen.

The ITD and ILD determine the cone of confusion and the spectral cues are used for localization within the cone of confusion.

Due to psychophysical experiments that show decreased azimuthal acuity in the mid-frequency range where neither cue is strong, and from the inability of human observers to direct ITDs in the ongoing fine structure of high frequency pure tones, the classic “duplex theory of sound localization” has been supported. It is important to note that the “duplex theory” holds for pure tones.

Simple stimuli such as pure tones, clicks and broadband noise are used to perform psychoacoustical studies on binaural detection.

3.4.1 Interaural Time Diference (ITD)

Interaural Time Diference is the major cue for localizing the azimuthal position of the sound source (Fitzpatrick Douglas). The ITD occurs because the distance between the two ears creates a sound wave which reaches the ipsilateral ear earlier that the contralateral, which is caused mostly by the shading effect of the head. Psychoacoustic experiments show that ITD dominate the high frequencies (<1.5 kHz) (Polley R. Liu 2006). ITD can be detected in the waveform of the signals or in the signals envelopes. The first case is relevant only for frequencies below 1.6 kHz. Detection of time differences in the envelopes is able for the whole frequency range above 100 Hz.

Interaural Time Differences can be mathematically modelled in a comparably simple way. Various models have been analyzed regarding the evaluation the ITD, caused by a plane waveform from certain directions. One of the simplest, let us say the simplest one, is the model which is only valid for acoustically transparent frequencies. In that case, diffractions around the head occur. The main dominant models for processing ITDs are based on the Jeffress model predicting neurons that fire maximally at a common ITD across their responsive frequency range (Jeffress 1948). ITD localization is intensively studied in biology.

3.4.2 Interaural Level Diference (ILD)

A significant role on human sound localization is the difference in level between the two ears, so named as Interaural Level Differences (ILD) (Blauert J. (1997). “Spatial.). In accordance with the famous duplex theory Interaural Level Differences (ILD), are the dominant cues for high-pass filtered stimuli, which are caused by the attenuation of sounds on the contralateral ear by the human head and the sound pressure amplitude decreasing with the distance of a sound source.

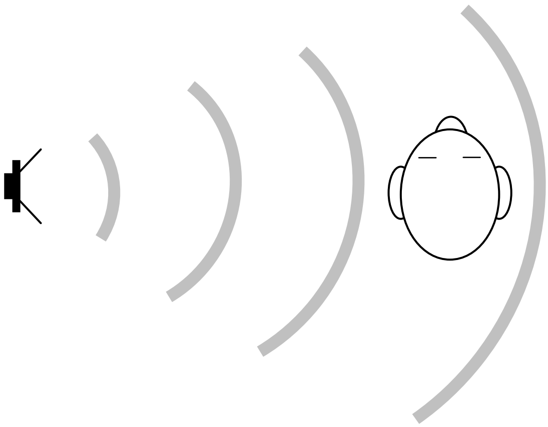


Figure 3.2 Iteraural Level Diference

The ILD constitutes the basis of the oldest theory of directional hearing; “intensity-difference theory” of directional hearing (Blauet, 1997 Spatial Hearing...), (Stanley Acoustic localization...).

The Interaural Level Differences can be considered as a function of the distance for very close sound sources. For source distances of r<1m, the distance between the two ears can not be neglected any more and the 1/r law becomes relevant. Brungart (1999 (Audit local of nearby sour HRTF)) indicates that ILD substantially increases for lateral sources at distances below 1m, even at low frequencies where ILD is usually small.

3.4.3 Interaural Intensity Difference

For localizing high-frequency sounds (above 1500 Hz) interaural intensity differences are used by the human brain. The Interaural Intensity

Diference represents the time integrated measure of stimulus level (Polley R. Liu 2006).

For any sound, there is a direct relationship between the direction from which the sound comes and the extent at which the intensity of the sound at the two ears differs. If the sound comes directly from the right, the sound will be lower in intensity in the left ear; if it comes from the front, the intensity at the two ears is the same; when the sound is coming from intermediate directions, there are intermediate intensity differences.

Intensity differences between the ears can result from two factors: differences in the distance the sound must travel to the two ears and differences in the degree at which the head casts a sound shadow. The greater the sound shadow cast by the head, the greater the level difference between the ears. The extent of the sound shadow cast by the head depends on the frequency of the sound. Low-frequency sounds have a wavelength that is long in comparison with the size of the head. The sound therefore bends very well around the head and there is very little sound shadow cast. In contrast, high-frequency sounds have a wavelength that is short in comparison with the dimensions of the head. This means that the head casts a significant sound shadow.

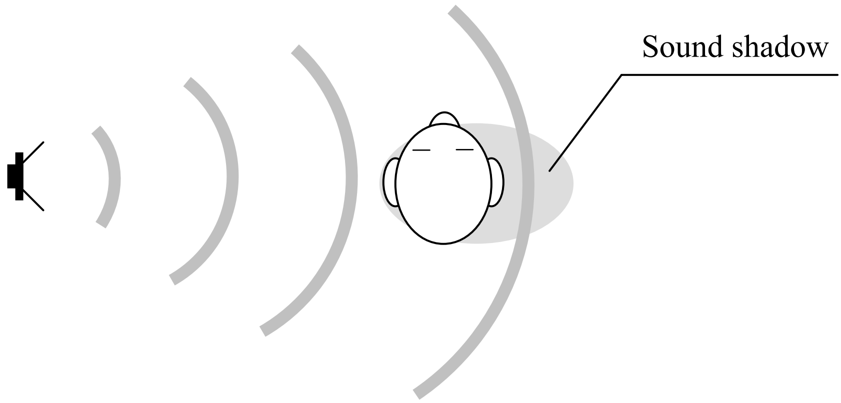


Figure 3.3 Interaural Intensity Difference

3.4.4 Cone of Confusion

It is well known that the interaural time differences and interaural level differences are important cues for sound localization. However, they do not specify exactly the direction from which the sound comes in a tree-dimensional space. Let us see for example, a sound in a median, horizontal plane; it will produce both an interaural time difference and an interaural level difference equal to 0. In that case, ITD and ILD will not be able to perceive the sound location, is the sound directly in the front, directly above etc. Thus, a cone of confusion describes locations in space where the binaural cues are identical; it is roughly symmetrically arranged around the interaural axis (Leung J 2004). Location of such cone of confusion may produce similar interaural level differences.

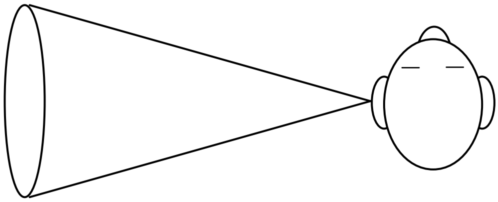


Figure 3.4 Cone of confusion

3.4.5 Precedence effect

The "precedence effect" refers to a group of phenomena that are thought to be involved in resolving competition for perception and localization between a direct sound and a reflection.

The precedence effect is thought to discount the reflected sounds in the computation of location, so that a listener perceives the source near its true location. According to most auditory theories, the precedence effect is mediated by binaural differences (Litovsky, 1997).

The standard theoretical model for this effect is an extension of the binaural model for localization, a neural coincidence detector that operates on the difference in arrival time of signals at left and right ears (Jeffress, 1948)

3.4.6 Reverberation effect

Reverberation accompanies any sound emerged in a natural acoustic environment. It results from reflection of any sound wave from the obstacles and its return to the point of listening. In a reverberant environment, sounds reach the ears through several paths. Although the direct sound is followed by multiple reflections, which would be audible in isolation, the first-arriving wavefront dominates many aspects of perception.

Normally, direct signal, which has the highest intensity, comes first. It is followed by early, or primary, reflections from walls, floor, ceiling which have lower intensity which depends on a distance covered and material properties. They are followed by secondary and multiple subsequent reflections with rapidly decreasing intensity. In a real situation sound impulses are lengthier than the time of arrival of first reflections. That is why reverberation is superimposed on a source sound.

3.4.7 Cross-correlation model

Most of ITD extraction models are based on Jeffers cross-correlation model and the cross-correlation implementation. The Jeffers model discusses on a network of coincidence detectors. Each one of the detectors is turned to a specific narrow band of frequencies and a specific ITD provides a natural mechanism for the computation of a set of narrow/band cross-correlation functions. The cross-correlation functions for the interval [0, T] are defined as:

$$R_{LR}(\tau) = \int_0^T x_L(t)x_R(t - \tau)dt \tag{11}$$

Where $x_L(t)$ and $x_R(t)$ are the left and right peripherally filtered signals.

Some of cross-correlation models include the combination of the evaluation of ILDs into the same structure.

Lindemann introduced one of the earliest models of lateral position based on cross-correlation, known as the triplex theory of hearing (Lindemann, 1959). Lindemann model includes monaural processors that are involved in ILD processing (Lindemann, 1986a). In this theory, Lindemann tries to formulate how stimulus waveform can proceed to provide the necessary information for understanding the hearing perceptions and its abilities. The Stern and Colburr model can explain the stimulus-dependence in the weighting of conflicting ITDs and ILDs (Stern and Colburr, 1978).

4. CHAPTER IV: SOUND SOURCE LOCALIZATION TESTS

CHAPTER IV: SOUND SOURCE LOCALIZATION TESTS

4.0. Summary

This chapter presents several experiments on sound source localization. Two experiments are developed using monaural clicks in order to verify the influence of the Inter-click interval on sound localization accuracy. In the first experiment, the localization of the static sound source was studied; the saltation perception on the inter-click presence was also analyzed. The experiment is based on monaural click presented at different inter-click intervals (ICI), from 10ms to 100ms. Two types of sounds single click and train of clicks are generated and thereafter tested at different inter-click intervals. At short inter-click intervals, the clicks were perceived as a blur of clicks having a buzzy quality. Moreover, it was proven that the accurateness in the response improves with the increase of the length of ICI.

The present results imply the usefulness of the inter-click interval in estimating the perceptual accuracy. An important benefit of this task is that this enables a careful examination of the sound source perception threshold. This allows detecting, localizing and dividing with a high accuracy the sounds in the environment.

In the second test, an analysis of moving sound source localization via headphones is presented. Also, the influence of the inter-click interval on this localization is studied. The experimental sound consisted of a short delta sound of 5ms, generated for the horizontal frontal plane, for distances from 0,5m to 5m and azimuth of 32º to both left and right sides, relative to the middle line of the listener head, which were convolved with individual HRTFs. The results indicate that the best accurate localization was achieved for the ICI of 150ms. Comparing the localization accuracy in distance and azimuth, it is deduced that the best results have been achieved for azimuth. The results show that the listeners are able to extract accurately the distance and direction of the moving sound for higher inter-click intervals.

4.1 Perception of the static sound source position

4.1.1 Sound sample

Sound source positions used for stimulus presentation in this chapter were generated for a horizontal frontal plane. A click of 5ms duration was generated with Above Audition software. Figure 4.1 is a picture of the considered click in Above Audition. The *x* axis represents the time scale, in seconds; the sound speed, in our case is 29,97 fps.

In the first case, the generated click with duration of 5ms was used as spatial sound and in the second case; the click was multiplied by six, becoming a train of clicks with duration of 30ms.

The click has been spatialized by using Head Related Transfer Functions (HRTFs). It is known that the HRTFs are very important for sound localization, because they express the sound pressure at the listener eardrum over the whole frequency range (see Figure 4.2). In the present study, the HRTFs were generated at 80dB at a frequency of 44100 Hz and processed by a computer for the frontal plane, for a distance of 2 m, with azimuth of 64º (32º at the left side of the user and 32º at the right side of the user).

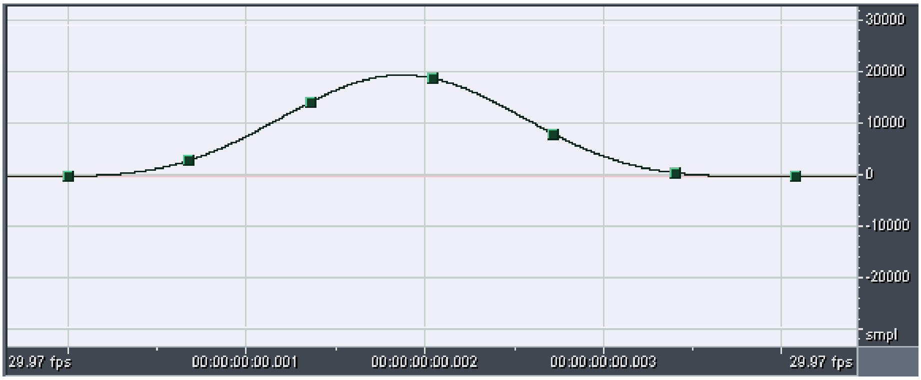


Figure 4.1 Click wave form with duration of 5ms. The x axis represents the time scale in seconds where the y axes represent the sound sample. The signal speed is 29,97 fps.

In the experiments the spatial clicks were presented randomly in pairs Left-Right and Right-Left, delivered using Matlab version 7.0, on an Acer laptop computer.

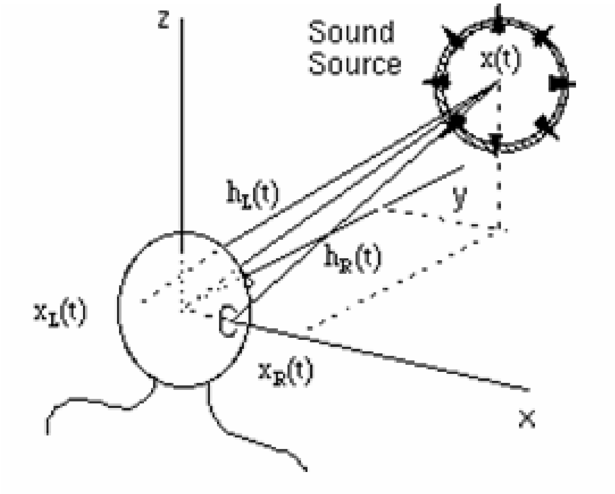


Figure 4.2 HRTF coordinates presentation, where the $h_L(t)$ and $h_R(t)$ represent, respectively, the head-related impulse response HRIR at the eardrum for the sound source $x(t)$ at each ear, left $x_L(t)$ and right $x_R(t)$. The $x_L(t)$ and $x_R(t)$ could be calculated using the convolution integral $x_L(t) = \int h_L(\tau)\, x(t-\tau)\, d\tau$ and $x_R(t) = \int h_R(\tau)\, x(t-\tau)\, d\tau$, where τ is the delay.

4.1.2 Test participants

Ten volunteers, 4 females and 6 males, age range 27-40 years, average 33,5 participate in this experiment. Each subject reported to have normal hearing, they did not reported any hearing deficiencies. All of them were supposed to other acoustical experiments with computer and acoustical mobility devices.

4.1.3 Procedure

The experiment was carried out in a single session. The session consisted of two runs, one for a single click and one for a train of clicks. Each run was based on six sounds. Fig.3 shows the schematic presentation of the sound: a) shows the monaural click in which, the click comes from (Left) L→R (Right) and R→L, with randomly varying ICIs; b) shows the train of clicks, where the presentation procedure is the same as for the single click, the sound come from L→R and R→L, with randomly varying ICIs. Different interclick intervals (ICI), from 10 ms to 100 ms were used (10ms, 12ms, 25ms, 50ms and 100ms).

Localization test were carried out in a chamber of 4,8m x 2,5m x 12m, where external sounds were present.

Since the experiments described in this chapter were focused on examining the perception in human listeners, it was important to be able to measure spatial capabilities in an accurate and objective way. For the localization test, subject localized auditory clicks presented in the headphones, telling the direction of the listened sound. In both cases the experiment begins with various exercises where the subjects are able to hear the clicks and train of clicks, separately, firstly the left one and afterwards the right one, continuing with the six sounds delivered by the program randomly. Afterwards the subject completed the all six sounds, the new exercises were presented of the combination Left-Right and Right-Left.

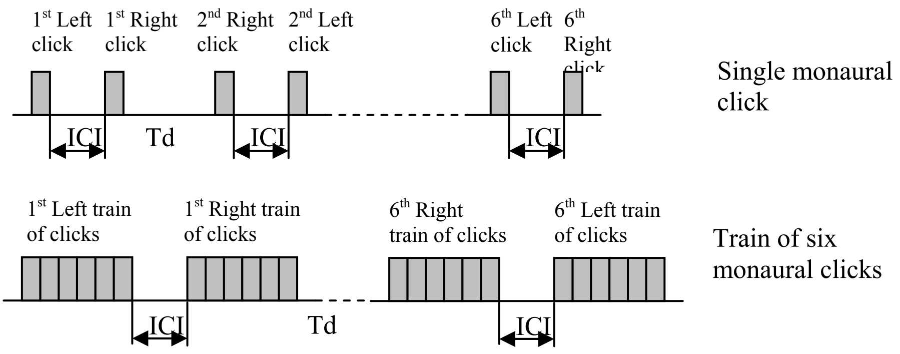


Figure 4.3 Schematic presentation of the sound. In both situations the click is of 5ms. In the first case, the click has been listened at the different interclick intervals ICI separated by a decision time Td. In the second case, the click has been substituted by a train of six clicks.

For the localization tests, listeners were sit comfortably in a chair in front of a computer. Before starting the test, the listeners received written and oral instructions and explanations of the procedure. They were asked to pay especial attention and to be concentrated on the experiment.

Before localization experiments, subjects had a training protocol to become familiar with the localization. This protocol included the speech pointing techniques, which requires that the subject verbally informs the evaluator about the perceived localization of a sound. During the experiment, since the subject had not access to the computer screen, the tendency of capturing the sound with the eyes was eliminated.

During the test, the subjects were supposed to listen through the headphones, model HD 201, twelve pairs of sounds; six pairs of single clicks and six pairs of trains of clicks Left-Right and Right-Left at different ICIs, from 100 ms to 10 ms in a decreasing succession.

The sounds were delivered in a random position. The sound used in the experiment was the same sound used in the testing procedure. The sound duration was brief enough, so that listener could not make head movements during the sound presentation. Between each two consecutive pair of clicks, the decision time Td was computed; this was the time needed for evaluating the sound (see Figure 4.3).

The subjects were asked what they listened, the number and the provenience of the listened sound and also if there was any difference between them. The subjects where allowed to repeat them, if necessary, after they had evaluated the perceived position for each click, classifying them as ‘Left’, ‘Right’ or possible ‘Centre’. Once the subject had selected a response, a next pair of clicks was presented. Each trial lasted approximately 2 min. The average time per subject for all experiment was around 35 min

Some distraction cues as: environmental noises, draw away seeing or hearing someone- since the subject remained with opened eyes- influenced on the experimental sound source perception and results. Because of this reason, the subjects were allowed to make judgments about the source location independently.

The results were collected by the evaluator and introduced manually into a previously prepared table. After the test, localization performances were examined using the analyses described in the following section.

4.1.4 Results

The results from the experiment were collected for data analysis. Localization performances summary statistics for each subject are listed in Table 1. The graphical user interface was generated by Excel in linear standard model.

Subject response was plotted in relation to the Inter-click Interval.

The main data for all subjects is presented in Figure 4.4 with an error of 5%.

The perception of the single and train of clicks and the perceived position of the sound pairs Left-Right and Right-Left were analyzed. Both factors as well as the interaction with the ICIs were significant.

Fig. 4 shows that the perception of the sound source position decreases when ICIs does. For avoiding errors, the tests results were registered up to an ICI of 10ms. Because ICI was enough short, the clicks were perceived as a single entity moving from one ear to another or from one ear to the centre having a buzzing quality.

In the case of the single pair of clicks at ICI of 12ms, because the length of the clicks and the length of the ICI were too short, the subjects could not distinguish clearly the clicks corresponding to the pairs Left-Right and Right-Left.

When comparing the perception of the single clicks with the perception of the train of clicks Figure 4.4 a), a great continuity of the sound position across almost the entire range of ICIs was detected. In other words, the perception of the sound position was stronger for the train of clicks. This effect may be a result of the better localization associated with the sound.

For ICIs between 25 and 10ms, the subjects perceive the Right-Left pair of sounds with a higher precision than that of pairs Left-Right for single click and train of clicks.

In other case, for ICIs of 50ms, the perception of the pair of single clicks Right-Left is higher than the perception of the pair Left-Right. In the case of the train of clicks, the perception results are equivalent for both pairs Left-Right and Right-Left.

Table 4.1 Localization performance summary statistics for all subjects (P1-P9) in frontal field. The percentage of the perception experiment is calculated on the basis of the six delivered sounds.

Click perception in %			**Train of clicks perception in %**		
interclick ms	***Azimuth -30º***	***azimuth 30º***	***interclick ms***	***Azimuth -30º***	***azimuth 30º***
100	100%	100%	100	100%	100%
50	90%	86%	50	100%	100%
25	80%	90%	25	88%	96%
12	83%	95%	12	76%	79%
10	88%	86%	10	75%	86%
8	100%	95%	8	100%	96%
6	100%	95%	6	85%	93%
5	100%	92%	5	100%	95%
1	100%	100%	1	100%	100%

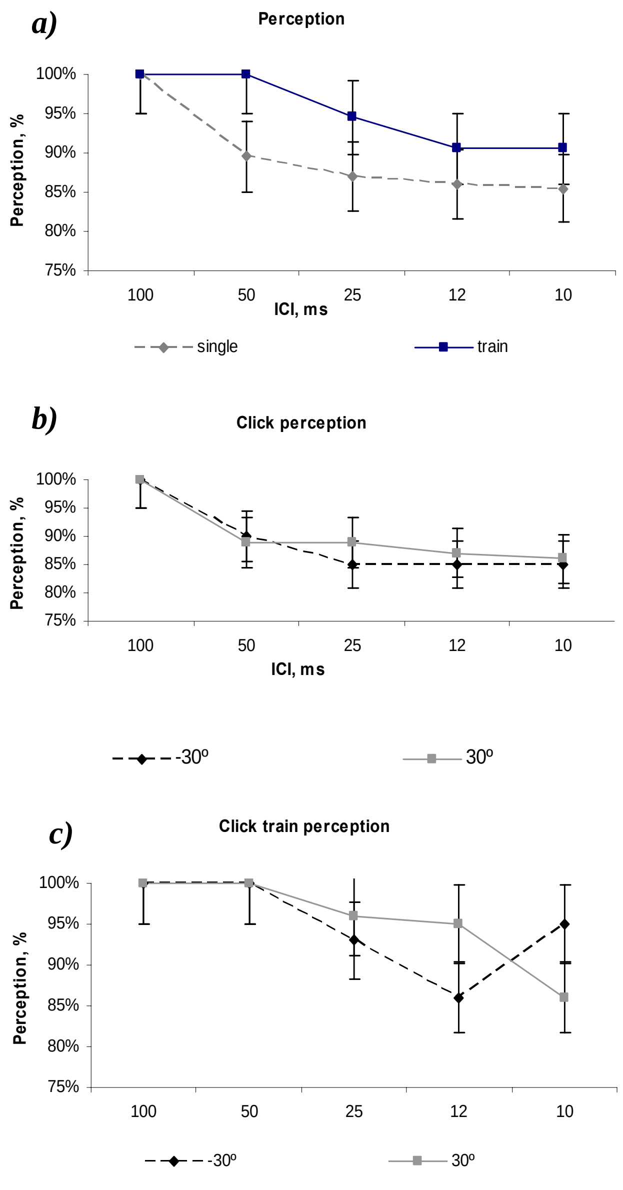


Figure 4.4 Mean estimation of the click location: a) Represents the perception of the single click and the perception of the train of clicks at 0º (center) at different ICIs; b) shows the click perception at -30º (left side) and +30º (right side); c) corresponds to the train of click perception at -30º (left side) and +30º (right side)

4.1.5 Conclusion

When trying to explain the sound source perception threshold, we perceive the perception of the saltation illusion. With shorter ICIs, a blur of clicks were perceived, in contrast with the individual clicks at longer ICIs. As the psychologist Gestalt noted, the perceptual system scrambles for the simplest interpretation of the complex stimuli presented in the real world. Therefore, the studies were based on analyzing and proving that, grouping the clicks, the sound source is better perceived and localized.

For longer ICIs, this procedure is not so important, since each click can be identified and localized.

The present results demonstrate the usefulness of the inter-click interval in estimating the perceptual accuracy. A possible benefit of this task is enabling a careful examination of the sound source perception threshold. This allows detecting, localizing and dividing with high accuracy the sounds in the environment.

4.2 The influence of the inter-click interval on moving sound source localization tests

4.2.1 Introduction

Humans have remarkable ability to perceive their surrounding through hearing. They are able to detect, identify and localize the sound source around them, to roughly estimate the direction and distance of the sound source, the static or moving sounds and the presence of an obstacle or a wall.

Sound source localization has been studied during many years (Brungart et al., 1999). Lord Rayleigh in his "duplex theory" presented the foundations of the modern research on sound localization (Stutt, 1907), introducing the basic mechanisms of localization. Blauert defined the localization as "the law or rule by which the location of an auditory event (e.g., its direction and distance) is related to a specific attribute or attributes of a sound event" (Blauert J., 1997). Acoustical cues (interaural cues such as Interaural Time Difference ITD and Interaural Level Diference ILD), torso and pinnae contribute to the human sound localization (Brungart et al., 1999), (Bruce H., 1959). Beside that, the Head Related Transfer Functions (HRTFs) are crucial for sound source localization (Kim et al., 2001). According to Blauert (1997), HRTFs represent the transfer characteristics of the sound source in a free field to the listener external ear.

Beside the localization of a static sound source, the moving sound localization plays an important role in the human life [Al'tman et al., 2005]. In the case of a moving source, changes in the sound properties appear due to the influence of the sound source speed or due to the speed of the used program for sound emission.

Several experiments have been carried out on static sound localization using headphones (Wenzel et al., 1993), (Blauert, 1997) but few for moving sound source localization. In the case of localization via headphones, the sounds are localized inside the head (Junius et al., 2007). This is known as "lateralization". Previous studies (Hartmann and Wittenberg, 1996) showed that sound externalization via headphones can be achieved using individual HRTFs, which help the listener to localize the sound out in space (Kulkani et al., 1998), (Versenyi, 2007). Great results have been achieved with the individual HRTFs, which are artificially generated and measured on a dummy head or taken from another listener. Due to those HRTFs, the convolved sounds are localized as real sounds (Kistler et al., 1996), (Wenzel, 1992).

Several experiments evaluating the sound localization have been carried out recently. In the first of these experiments (Dunai et al., 2009) the localization of the position of a single sound and a train of sounds was carried out for different inter-click intervals. The initial sound was a monaural delta sound of 5ms processed by HRTFs filter. The ICIs were varying from 10ms to 100ms. The listeners were asked to inform what they listened, the number and the provenience of the listened sound and also if there was any difference between them, evaluating the perceived position of the sound (“Left”, “Right” or “Centre”). It was proven that the accurateness in the response improves with the increase of the length of ICI. Moreover, the train of clicks was localized better than the single click due to the longer time to listen and perceive the sound provenience.

In the second study (Dunai et al., 2009), the real object localization based on sensory system and acoustical signals was carried out via a cognitive aid system for blind people (CASBliP). In this research, the blind users were walking along a 14m labyrinth based on four pairs of soft columns should localize the columns and avoid them. The average time of sound externalization and object detection was 3,59min. The device showed no definitive results due to the acoustical signal speed, which required improvements.

The goal of this study is to analyze how localization of a moving source is influenced by the inter-click interval, and how the listeners localize the moving sounds through the headphones.

4.2.2 Subjects

Nine young subjects with ages between 25 and 30 years and different gender, all of them had normal vision and hearing abilities, were involved in the experiments. All participants had normal distance estimation and good hearing abilities. They demonstrate a correct perception of the sounds via headphones. The subjects were identified by a number P1-P9.

All subjects participated in previous auditory experiments in the laboratory. Each participant received a description of what was expected of him/her and about all procedure. All participants passed the localization training and tests described below.

4.2.3 *Stimuli*

A delta sound (click) of 2048 samples and sampling rate of 44.100 Hz was used. To obtain the spatial sounds, the delta sound was convolved with Head-Related Transfer Function (HRTF) filter measured for each 1º in azimuth (for 32º left and 32º right side of the user) at each 1cm in distance (See Figure 4.5). The distance range for the acoustical module covers from 0,5m to 5m, an azimuth of 64º, and 64 sounding pixels per image at 2 frames per second.

Recording of directional transfer functions were carried out in an anechoic chamber (Figure 4.6)

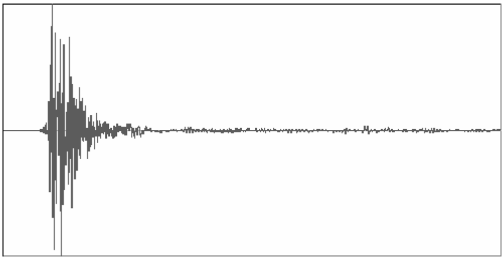


Figure 4.5 HRTF wave form of 22050Hz sampling rate, and the length of 46ms of 8192 bit. In the x axis, the stimulus sample is represented.

The HRTFs measurements system is based on a robotic and acquisition system. The robotic system consists of an automated robotic arm, which includes a loudspeaker, and a rotating chair on an anechoic chamber. A manikin was seated in the chair with a pair of miniature microphones in the ears. In order to measure the transfer function from loudspeaker-microphone as well as for headphone-microphone, the impulse response using Maximum Length Binary Sequence (MLBS) was used. The impulse response was obtained by taking the measured system output circular cross-correlation with the MLBS sequence. The impulse response is given by:

$$h(n) = \Omega_{sy}(n) = s(n)\Phi y(n) = \frac{1}{L+1}\sum_{k=0}^{L-1} s(k)\cdot y(n+k) \qquad (12)$$

Figure 4.6 The anechoic chamber. The walls and floor of the chamber are lined with sound absorbing wedges. The whole robotic system can be moved to place the loudspeaker at any location in radius of 5 meters. A chair for manikin is fixed in a platform.

where *y(n)* is the system output, *s(n)* is the MLBS and Φ represents the circular cross-correlation.

Because the direct implementation of the equation (1) requires a long processing time, we used the equivalent operation of the cross-correlation, convolution passing to the frequency domain. In that case, the convolution is a vectorial multiplication:

$$a(n)\,\Phi\, b(n) = \frac{1}{L+1} a(-n) * b(n) \tag{13}$$

In order to reduce the computational time, the Fast Hadamard Transform (FHT) has been used. In that case the impulse response is given by:

$$h(n) = \frac{1}{(L+1)s[0]} P_2 \left\langle S_2 \left\{ H_{L+1} \left[S_1 \left(P_1 y(n) \right) \right] \right\} \right\rangle \tag{14}$$

where P are the permutation matrices, S – the redimention matrices and y H_{L+1} is the Hadamard matrix of L+1 degree.

The output signals (the HRTF) are sampled at 22050Hz and a length of 46ms (8192 bit).

The HRTFs were measured for the horizontal frontal plane at the ear level from 0,5 to 5m in distance and in azimuth between 32º left and 32º right with respect to the centre of the listener head (measurements at every 1º). Figure 4.7 shows the graphical representation of the stimuli processing

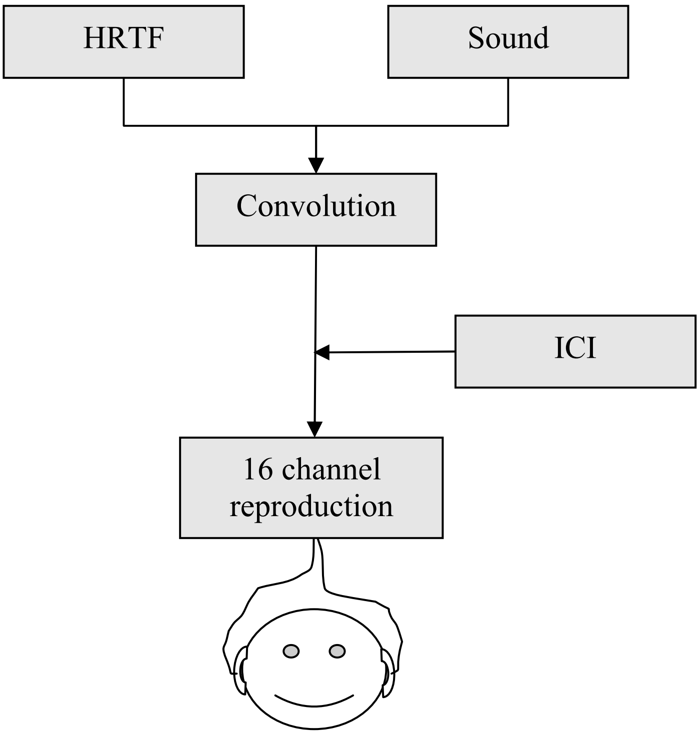


Figure 4.7 Method for sound processing and reproduction

4.2.4 Equipment

For the sound generation and processing, a Huron system with 80 analogue outputs, eight analogue inputs and eight DSPs 56002, and a computer for off-line sound processing was used.

For the experimental test, SENNHEISER headphones model HD 201 were used to deliver the acoustical information. The model was selected because it has a good stereo sound and it attenuates the ambient noises; minimum interference with external sounds is desirable in order to obtain best acoustical results.

MATLAB 7.0 was used as experimental software. The resultant graphical sound trajectory for each experiment was displayed on a separate window and saved for off-line processing.

All experiments were run in an ACER Aspire 5610 computer.

4.2.5 Procedure

The goal of the experiments is to analyze the localization of a moving sound source via headphones and to see how the inter-click interval (ICI) influences the sound localization quality.

The comparison between the localization performances enables to evaluate the importance of the inter-click interval parameter for its use in sound localization and acoustical navigation systems.

The movement of the sound source was achieved by switching the convolved sound for a frontal plane at the eyes level at increasing distances from 0,5 to 5m (1 cm increase) and for azimuth between 32º right and 32º left (1º increase) with respect the middle of the head. The sounds were delivered for five inter-click intervals [200ms, 150ms, 100ms, 75ms and 50ms]. Figure 5.4 shows one of the trajectories the sound was running. Four different trajectories were created. The delivered trajectory was selected randomly by the computer when the experiment starts.

Before starting the experiment, the training exercises were carried out; the objective and the procedure of the experiment were explained to each individual participant. One sound was delivered for all five ICIs, where the participants were able to see graphically the listened sound trajectory (See Figure 4.8). In order to proceed with the test and experiment, the participants were asked to seat comfortably in the chair in front of a computer (See Figure 4.9). After reading and testing the training exercises, the participants were supposed to carry out the experiment.

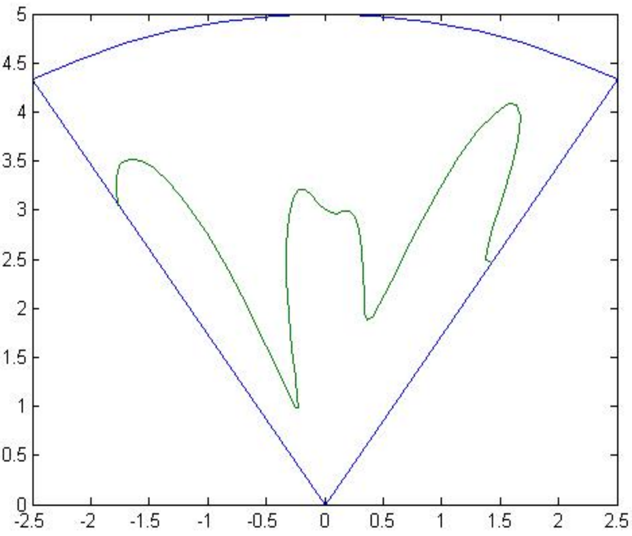


Figure 4.8 Sound trajectory example, direction from left to right. The x axis represents the azimuth where the 0 is the centre of the head, which is 0º. The -2.5 is the -32º at left side of the head and 2.5 respectively is 32º at the right side of the head. The y axis represents the distance from 0 to 5m

Figure 4.9 Experimental scenario. The user is seated on a chair in front of a computer. Hearing the sound through headphones he should draw the perceived sound trajectory in the paper

A sound at a specific ICI was delivered by the computer via headphones. During the experiment, the participants were free to move. Nevertheless, they were required to move the less possible and to be concentrated on the sound, in order to create a plane of the sound route in the imagination. The test was performed both with open eyes and with closed eyes depending on the participant wishes. In the case of the closed eyes, there was a limitation of effects of the visual inputs. Due to this, the participant achieved a better interpretation of the trajectory image.

The participants were asked to carefully hear the sound and draw the listened trajectory in a paper. They were allowed to repeat the sound if it was necessary. All the participants asked to repeat the sound at least three times. Each participant was supposed to have five trials, one for each ICI. Only one sound trajectory was used per participant for all five ICIs. For all participants, the experiment started with the ICI of 200ms, decreasing it progressively up to 50ms.

After the experiment the participants commented the perceived sound trajectory and they compared the listened sound for each ICI.

4.2.6 Results

The moving sound source localization is an important factor for the navigation task improvement. The main variables analyzed in this paper were the moving sound source localization and the inter-click interval ICI [200, 150, 100, 75, and 50ms]. The study analyzes the interaction between these variables in measurements of distance and azimuth.

Generally, no significant differences on the results were registered between participants. However, great difference was found in the sound localization between higher and lowers inter-click intervals.

The minimum and maximum data of the distance and azimuth are presented in Table 4.2 where, the maximum displacement in distance is 1,26m for an ICI of 50ms and the minimum displacement was 0,42m for an ICI of 150ms, the maximum displacement in azimuth was 11,4º for an ICI of 50ms and the minimum 0,71º for an ICI of 150ms.

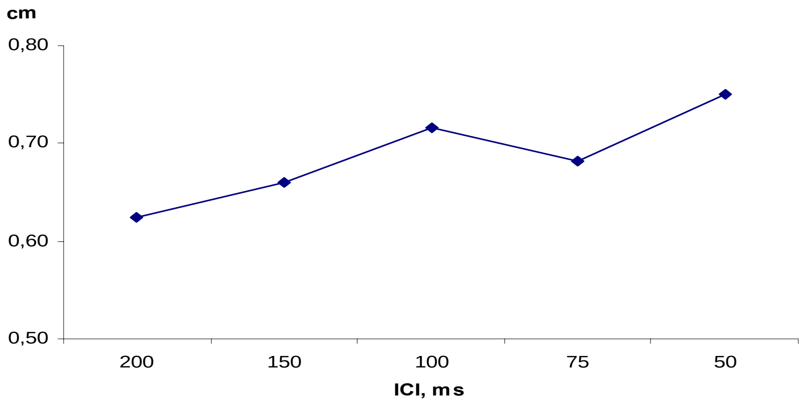
Table 4.2 Evaluation of the minimum and maximum displacement for all participants as a function of the inter-click interval

ICI, ms	distance, m **min**	**max**	Azimuth, º **min**	**max**
200	0,44	0,88	2,1	7
150	0,42	1	0,71	6
100	0,47	1,01	1	9,14
75	0,56	0,85	2,14	14
50	0,5	1,26	1,43	11,4

Average results of sound localization in azimuth and distance as a function of the inter-click interval are shown in Fig. 4.10. Best results have been achieved for greater ICIs, due to the time needed by the brain to perceive and process the received information. Because the time between two sounds is higher, the sound is perceived as jumping from one position to another from left to right in equal steps. For the ICI of 200ms, the sound was not perceived as a moving sound, but rather as a jumping sound from location to location. However, for the ICIs lower than 100ms the sound was perceived as a moving sound from the left to right, but there was enough difference between the original sound trajectory and the perceived one. The participants had great difficulties to perceive the exact distance and azimuth, because the sound was delivered too fast. Moreover, when the sound trajectory had multiple turning points on a small portion of the space, the participants perceived this portion as one turn-return way. Figure 4.10 represents a specific case, corresponding to one of the participants; it shows the moving sound localization at four ICIs. The red colour represents the listened sound trajectory drawn by the participant. The grey colour represents the real sound trajectory drawn by the computer. The *x* axis represents the azimuth where the 0 value is the centre of the head, the negative values are the values at the left side of the head, whereas the values at the right side of 0 represent the azimuth values at the right side of the human head. The -2.5 represents the 32º at left side of the head and 2.5 the 32º at the right side of the head. The *y* axis represents the distance from 0 to 5m.

In some cases, the participants perceived the sound trajectory as an approximate straight line when the inter-click interval was 50ms. Even repeating several times the experiment, the participants were confused regarding the localization of the moving sound. They commented "the sound moves too fast and I feel that it is running from left to right in a straight line". Despite listeners were not able to localize the moving sound source at lower inter-click intervals so well as they were able to localize the moving sound for greater inter-click intervals, they were able to judge about the sound position in azimuth and distance.

Various factors as drawing abilities (how the participants can accurately draw), sound interpretation (how the participants can interpret the heard sounds, by colours, by image etc.), the used hearing methods (with closed or opened eyes), the external noises, etc., influenced the experiment results. Despite all participants were informed about the use of one sound per participant for all ICIs, they draw the trajectories at different distances (see Figure 4.11).



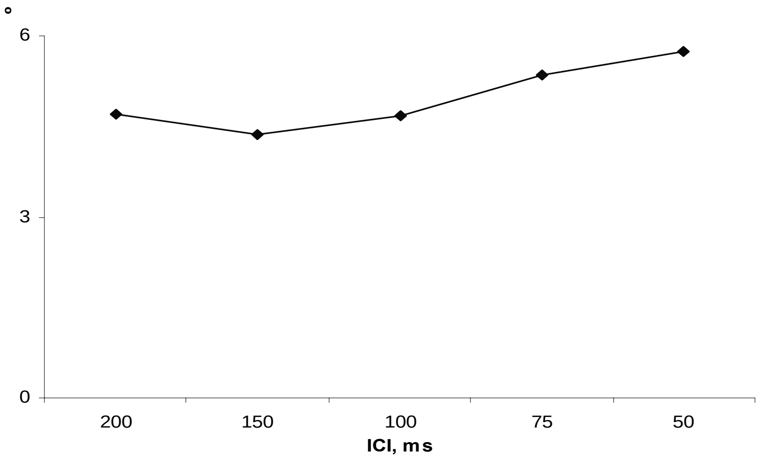


Figure 4.10 Average displacements in azimuth and distance for all participants

This error appears because of the participant drawing ability; it is not so easy to interpret graphically what is listened or the image the brain creates if there is not practice on that. For some of participants, great concentration and relaxation was required, to be able to correctly perceive the sounds.

Multiple observations on training sound trajectory were given to participants about how to perceive the sound and to be confident of their answer. Two participants were excluded from the main analysis due to the difficulties in localizing the sound. The participants experienced the moving sound localization as a straight line for all inter-click intervals.

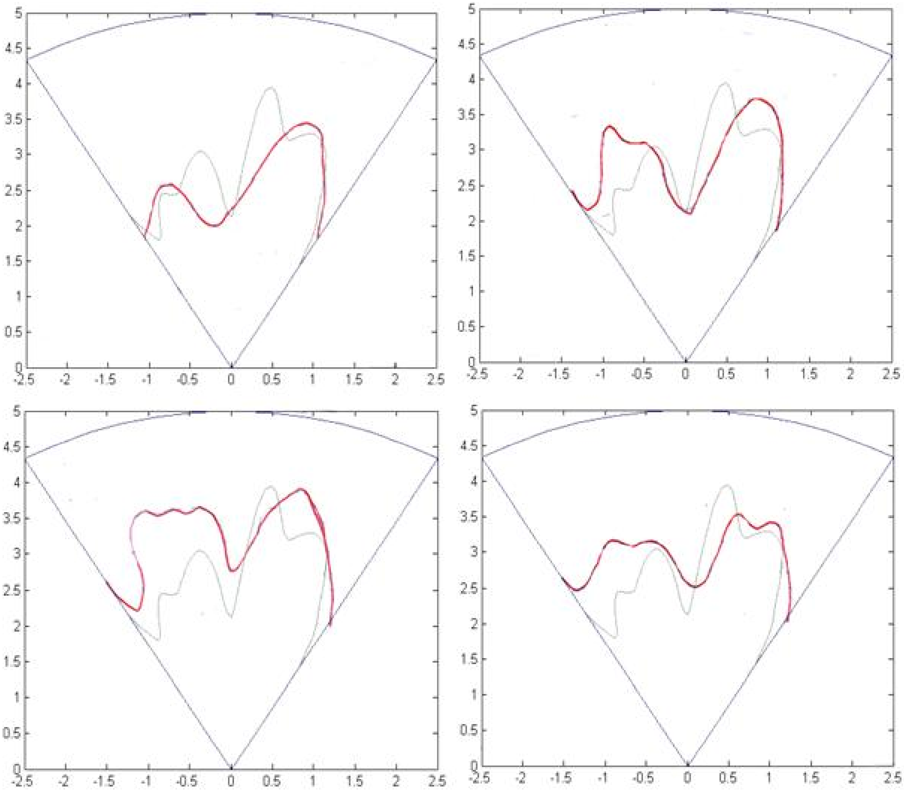


Figure 4.11 Sound trajectory for one participant for the ICIs of 50ms, 100ms, 150ms and 200ms. The red colour represents the heard sound trajectory drawn by the participant; the grey colour represents the real sound trajectory drawn by the computer. The *x* axes represent the azimuth, in which the 0 value is the centre of the head, the negative value are the values at the left side of the head and the values at the right side of 0 represent the azimuth values at the right

side of the human head. -2.5 represents the 32º at left side of the head and 2.5 respectively the 32º at the right side of the head. The *y* axis represents the distance from 0 to 5m.

4.2.7 Conclusions

The results showed that when the inter-click interval is higher, the moving sound is better localized, both in distance and in azimuth. The results indicate that the best accurate localization was achieved for the ICI of 150ms. The analysis indicates that for shorter ICIs, all participants needed additional times to repeat the same experiment. The study has shown that the localization of a moving sound source plays an important role in the human life when it is necessary to guide oneself in the environment.

4.3 General conclusions

In the present chapter two sets of experiments are described according to the examined spatial performance involving simple broad-band stimuli. Both experiments measured how well single and train of static and moving sounds are localized in laboratory conditions. These experiments demonstrated that sound source is essential for accurate three-dimensional localization. The approach was to present sounds overlapped in time in order to observe the performance in localization, in order to see how time delay between two sounds (ICI inter-click interval) influences on sound source localization. From the first experiment it was found that better localization performance was achieved for trains of sounds at an ICI of 100ms. If analyzing the localization results at the left and right side of the human head, it must mention that improved results were obtained at the left side for the single click and at the right side for the train of clicks. At short inter-click intervals, the train of clicks was perceived as a blur of clicks. At short inter-click intervals the single clicks was perceived as one click, there were not perceived the difference between the first click and the second one. In this case only the first click was perceived, the second click was perceived as a week eco. Moreover, the sound perception threshold was studied. In the second study the localization of a moving sound source both in distance and azimuth was analyzed. The results demonstrate that the best results were achieved for an inter-click interval ICI of 150ms. When comparing the localization accuracy in distance and azimuth, better results were obtained in azimuth. The maximum error in azimuth is of 11,4º at the ICI of 50ms. The disadvantages of the results at short ICI’s are due to that the total time of the sound run is very short, that prevent the user to perceive all the sound coordinates. Regarding the large ICI’s, the saltation from one click to another don not allows the user to make the connection between the two clicks. From this motive the user perceive the sounds as diffuse. Spatial cues such as interaural time difference ITD and interaural level difference ILD play an important role in spatial localization due to their attribution on the azimuthal sound localization. They arise due to the separation of the two ears, and provide information about the lateral position of the sound.

5. CHAPTER V: OBJECT DETECTION THROUGH ACOUSTICAL SIGNALS

CHAPTER V: OBJECT DETECTION THROUGH ACOUSTICAL SIGNALS

5.0. Summary

The chapter 5 describes the tests carried out for object detection with the cognitive aid system for blind people based on acoustical signals. In the first section of the chapter, preliminary tests on object detection and localization are analyzed in order to validate the system. The navigation and detection error and the sound externalization perception were studied. The experimental tests are based on the measurement requirements described in the Chapter 1.The protocols designed in the Chapter 2 are validated for the future experimental tests. Spatial signals were used to generate the cognitive task when the participants (blind users) were walking along a specific labyrinth. The information provided by the system indicated in real time the precise location of the obstacles both in distance and azimuth, by means of spatial signals. Thanks to the sensor precision and characteristics, such as those of the acoustical signal, significant results were obtained. After the first set of experiments the acoustical system has been improved, decreasing the inter-click interval and improving the sound source properties. The experiment consisted of two different tests; the Basic Learning Protocol, during which the participants became familiarized with the spatial signal, sound externalization and object recognition; the navigation along the labyrinth. The participants showed great results both in time and in number of hits; a minimum time of 0,41 min was achieved to complete the whole trajectory for the laboratory trials, whereas the number of hits which decreased to 0 for the second trajectory.

The results of the experiments based on acoustical signals and time of flight laser show that the system enables to control the localization of the real obstacles both in distance and in azimuth. The results have direct implication on human everyday navigation and guidance for visual impaired people.

Besides the environment data acquisition system, an important factor for the blind user is the representation of the acquired information, i.e., how to inform the blind user that the system detected an object in the front of his view.

5.1 Preliminary test I

5.1.1 Subjects

These tests were carried out on four users; three of them were blind, whereas the fourth had normal vision. One of the blind users (A) was a partially blind user, who was experienced in testing various types of electronic way-finding technologies. The second blind user (B) was a young blind man who lost his vision in 2003. The third one (C) was blind since birth. All participants had normal hearing abilities demonstrating correct perception of the spatialized sounds.

5.1.2 Procedure

Several tests were developed in order to verify the different features of the system. The tests were carried out on the basis of a learning protocol. The indoor navigation environment was a large hall with 15 m length and 10m width. The outdoor navigation environment was a park. The time interval between each set of spatialized sounds was 153ms.

The study consisted of two stages: (1) familiarization with the system functionality and spatialized sounds and (2) navigation in the real indoor and outdoor environment. At the beginning of the first phase, the subject received a short and concise explanation about the features of the system and how to manage it. A series of tasks were included regarding sensor acquisition, audio feedback and volume.

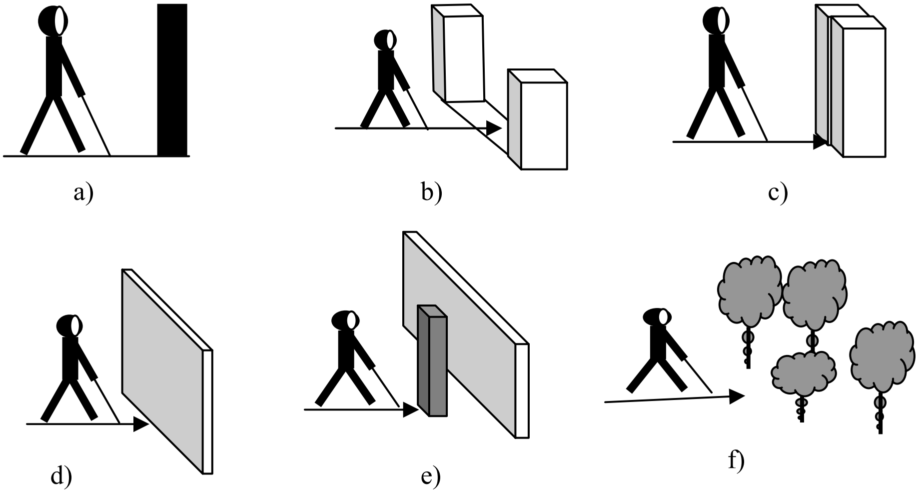


Figure 5.1 Experimental scenarios: a) Single column detection; b) Two columns detection and pass through them; c) A wide wall; d) A column detection in front of a wide wall and e) Outdoor experiment

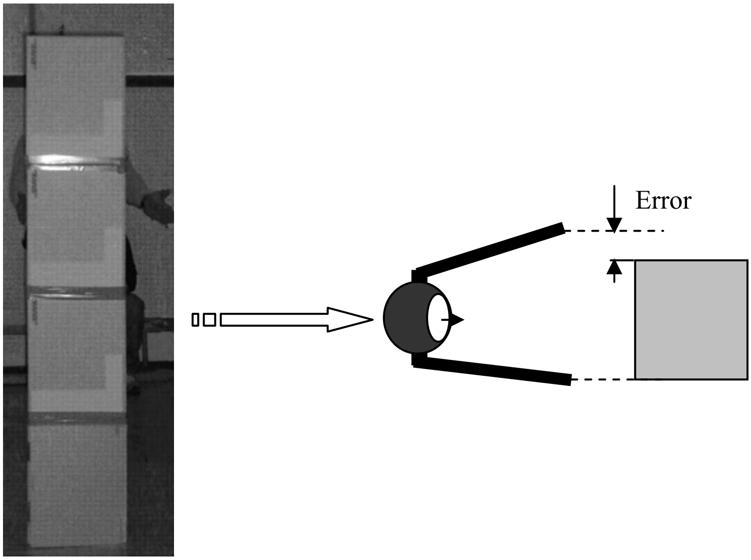


Figure 5.2 Distance and volume recognition representation

One of the objectives of the tests was the “externalisation” of the sound source. The users should perceive the sounds like coming from outside, from the objects themselves rather than being in the ear. In this way, the recognition of the location of the objects and the perception of their height, width and distance can be achieved (see Figure 5.2)

In the second stage, the exploration and recognition of the objects, while the user walked from a certain point towards the point in which the object was placed, were studied. During the walking task, the subjects should carry out a series of exercises consisting of overcoming several soft objects with different dimensions, which were arranged in a specific way. Some different situations were taken into account: a single column (Figure 5.1 (a)), a free passage between two columns (Figure 5.1 (b)), a wide wall (Figure 5.1 (c)), a single column in the front of a wall (Figure 5.1 (d), and finally, an outdoor area (Figure 5.1 (e)).

In all the situations, the subject starts the experimental test at distance of 3 meters far from the object, faced in a direction such that the distance between the wall and the user is longer than 5 m. Thus, neither the system will detect any object nor the subject will hear any sound from the system. The user starts looking around in order to explore the environment. Whenever he detects an object in his direction of view, he receives an external beeping sound through stereo headphones. In the case that the subject gets near the column (object), the sounds increase in intensity.

The intensity of the spatial sound is inversely proportional to the distance. At the same time, subject must comment what he listens and how he perceives the sounds using his own imagination.

In the case that there are more than one object (situations of two columns, Figure 5.1 b), a column situated in the front of a wall, Figure 5.1 d), and in the outdoor environment, Figure 5.1 e)), the subject listened some differences in the intensity of the sounds depending on the distance of each object. For the situation of the column situated in the front of a wall the subject will listen an intense sound sequence coming from the column and a secondary sound, as a background, coming from the wall.

In every experiment, the user is requested to indicate correctly the edges with his arms outstretched and gauge the distance and width of the gap (Figure 5.2).

5.1.3 Results and discussion

The data were collected in the five aforementioned scenarios from four subjects (three blind and one subject with normal vision). These data are shown in Figure 5.3. Figure 5.3 (a) indicates the times in which the users perceived the sounds for different trials (1,2 and 3). Figure 5.3 (b) shows the average times required by the user for the detection and location of the objects in trials 1,2 and 3. The figure indicates that the average time for sound source perception was quite similar for all users in each trial, except on subject C, who had difficulties in the learning of the system. The tests indicate that the error in distance perception is +/- 40cm, which means that the subjects were wrong with only one or two steps. Regarding the height and width, the error rate was +/-10 cm.

As it can be seen in Figure 5.3, the familiarization with the system functionality requires a short time (only few minutes). Due to the sound frame rate which is quite slow, it was difficult to walk with a normal rhythm. There were individual differences in performances: some participants had little difficulties with the guidance mode. In Figure 5.3, it can be seen that the subject C, who is total blind since birth, has some difficulties when externalizing the sounds and perceiving the object localization. For this user, additional time in the trainings was necessary. In general, participants performed very well on these trials. After some trials, the users were able to perceive the sound origin and to localize the objects in few seconds. The average time for the sound perception was 2,32 min and for the object detection 4,86 min. The total average time for completing these exercises was 3,59 minutes.

After the indoor trial sessions, additional tests were developed in outdoor environment. After a training day, the user was able to find the way outside quite easily. The exercise was certainly complex, since in order to go outside the user should pass through the door of the laboratory, through a corridor with many corners and finally to arrive outside. The subject was able to detect and gauge the size of obstacles such as a bicycle or a car, though she was not able to identify them.

Another situation was based on detecting the presence of other persons standing in the front or moving. In this situation, the user was able to gauge the distance at which other people were. The results suggest that with the used system it is possible to perceive the presence, the position and the dimensions of the detected object. This is indicating that the spatialized sounds are a promising solution as a part of the user interface of a blind navigation system. Performances were better than other models investigated before. The spatialized acoustical sounds have the additional advantage of consuming shorter time than conventional speech.

The present data enhance and expand the previous studies, which have demonstrated the utility of direct perceptual cues to navigation. It will be important for future work to compare the guidance models in more complex environments, such as train stations, supermarkets, steps etc. Another question for further research is the sound source localization and sound cues characteristics.

The present findings may have good applications in guiding navigation for the blind people. An advantage of the system is that could be integrated with other navigational systems such as GPS and other visual interfaces.

An important result of the Acoustic Interface is that ear-phones do not exclude real sound appearing from outside

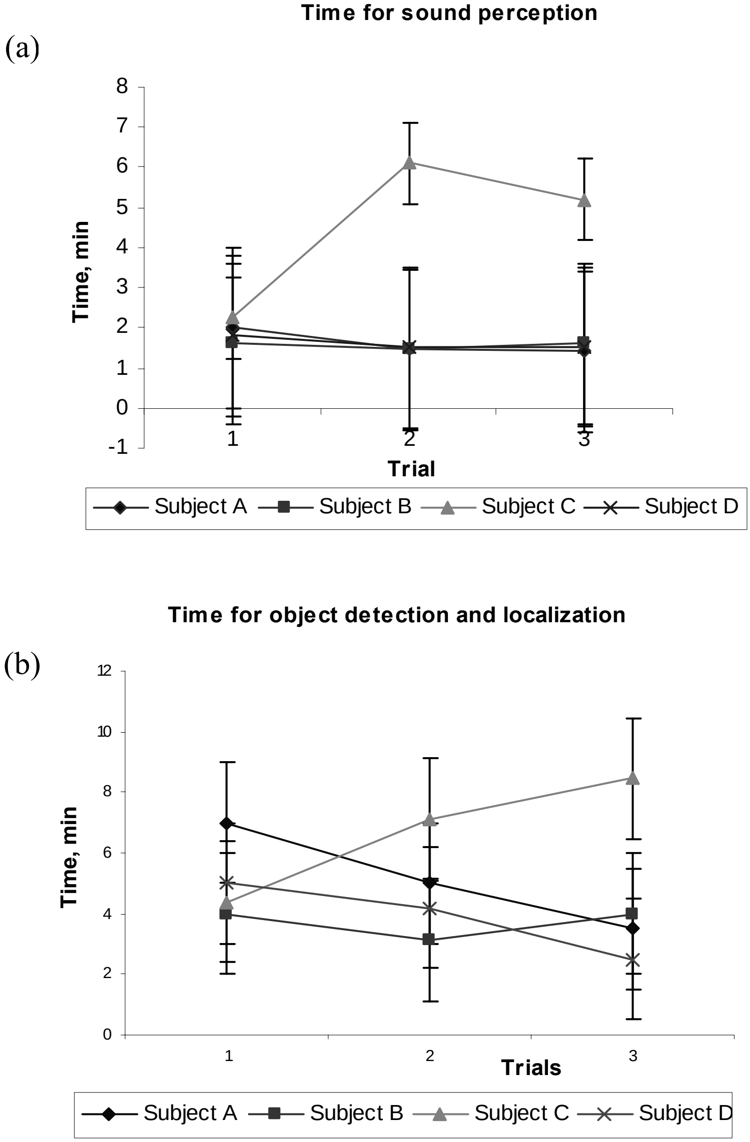


Figure 5.3 Results of preliminary tests on auditory localization; (a) Time for sound perception, (b) Time for object detection and location

.

5.1.4 Conclusions

Preliminary test I for object detection via acoustical signals show that the information generated by the Acoustic Prototype can guide the blind user safely around the environment. The device was able to detect with high precision the presence of the objects in the system area of view and to detect the object volume. During the tests, two main data were collected: the time for sound perception and time required for object localization. From the experiments, we can see that the average time for sound perception is 2,32min. This is the required time to learn to externalize the sounds, to perceive them as coming from the surrounding environment and not from the headphones. The average time for object detection is 4,86 min. Since the subjects were testing from the first time the system and they did not have any experience with the acoustic navigation systems mainly with the Acoustic Prototype, they had difficulties in becoming familiarized with the cognitive part of the system, to externalize the sounds and attribute them to the surrounding objects. Once again the experiments show the importance of training.

5.2 Preliminary test II

5.2.1 Procedure

After performing each practice with the Basic Learning Protocol, the research assistant and blind trainer were giving suggestions and answering the participant doubts. During the BLP tests, the time for sound externalization, object detection and error on distance and azimuth were registered.

Afterwards, when the participants got certain experience with all seven exercises, they were supposed to carry out the main test – localization of the obstacle and navigation through the labyrinth.

The participants were guided towards a free space, in which there were no obstacles in the front of view, in an area of 15m^2. After the device had been connected and prepared, the user was faced to the scenario. The participants were requested to navigate through a 14m long labyrinth, with four pairs of soft obstacles of 180cm height. The columns were placed with a horizontal separation of 2m and a distance of 2,5m between the last pair of columns and the wall (Figure 5.4). Two routes were prepared for the experiments. The first route had four turning points, including two 20º turns Left and Right and two 40º turn Right and Left. The second route had six turning points including two 25º turns Left and Right, two 40º turns Right and Left and finally two 45º turns Left and Right.

Firstly, the user was analyzing the environment, detecting the first pair of columns; he walked among them up to the next step where it was necessary to detect the next two pair of columns passing between them and so on…, until completing the whole labyrinth.

When the object was detected, through the headphones the participants listened spatialized sounds indicating the object presence, allowing him to rough it and pass beside. The sound used on the navigation test was the same as for the BLP, in order not to cause confusion on the blind user.

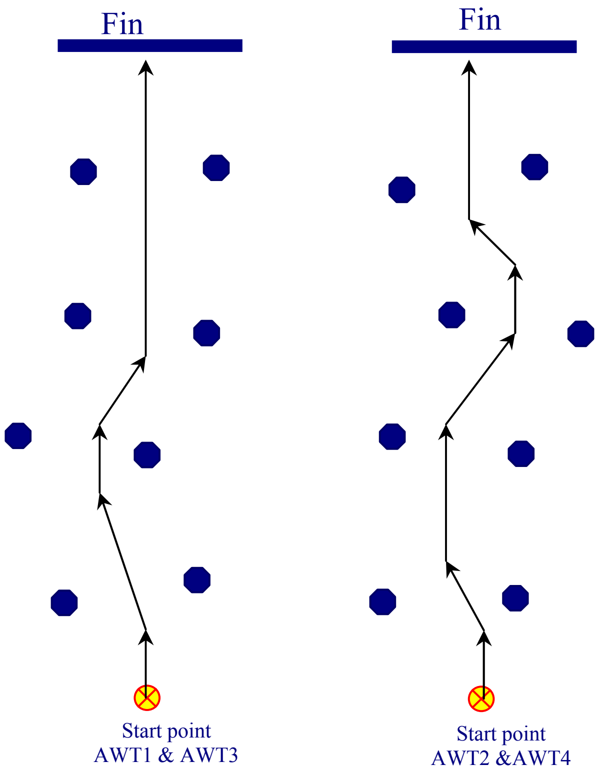


Figure 5.4 Experimental scenario

5.2.2 Results and discussion

The average time for all subjects was 3,6min per path; the maximum time was 7,31min and the minimum time spent to complete the whole trajectory was 0,41min. The average number of hits was 1,55 hits.

Figure 5.5 shows that the second group needed longer times and had higher number of hits (AWT3, AWT4 and NH3, NH4) due to the characteristics of the selected participants (5 blind participants with good orientation and five participants born blind with low orientation) who had difficulties with the transition from the white cane and felt unsafely; however, they on the second trajectory AWT4. This can be seen in Figure 5.6 which shows each group performance on the trials. Moreover, the number of hits for the second group was higher than for the first group. From Figure 5.5, it can be seen the improvement of the results both in time AWT4, with 1,37min, and in number of hits NH4, with 1,5 hits, for the second trajectory, where the participants were more relaxed and self-confident. The maximum number of hits was registered on the second group for the first trajectory, where one subject had 5 hits and the average number of hits was 3,2 hits. One of the ten participants of this group achieved 1 hit for the first trajectory and 0 hits for the second trajectory. Besides, a maximum time of 7,31min was registered on the second group on the first trajectory.

The best results both in time and in number of hits were recorded on the first group on the second trajectory, where five participants had 0 hits and other five participants had only one hit. Regarding the time, the minimum time spent to complete the whole trajectory was of 0,45 min, with a standard deviation from the Absolute Walking Time of 0,118 min.

Table 5.0 Navigation accuracy

First group

	TP1	TP2	TP3	TP4	TP5	TP6	TP7	TP8	TP9	TP10
AWT, min	2,45	6,58	4,21	6,45	3,18	6,01	4,15	4,32	7,31	7,02
AWT1, min	2,23	3,06	2,17	4,48	2,14	5,27	2,25	3,29	6,58	6,54
NH	2	4	1	3	3	5	4	3	4	3
NH2	2	1	0	2	2	3	2	1	2	2
AWT2, min	0,28	0,29	0,35	0,26	0,25	0,4	0,28	0,32	0,47	0,42

Second Group

	TP11	TP12	TP13	TP14	TP15	TP16	TP17	TP18	TP19	TP20
AWT, min	1,41	2,12	1,54	2,15	3,4	3,1	3,1	3,5	2,2	3,51
AWT1, min	5,54	2,25	0,45	2	3,27	1,51	3,39	3,4	4,35	3,2
NH	1	0	3	1	0	0	1	1	1	0
NH2	1	1	0	0	0	0	1	1	1	0

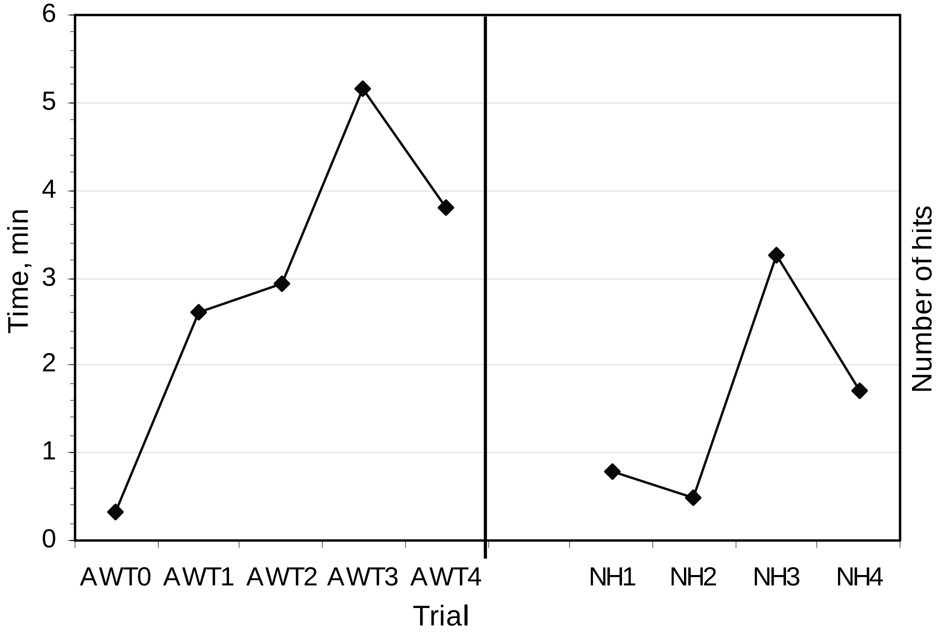


Figure 5.5 Time and number of hits for four experimental trials including the time with the white cane (AWT0)

Different variants for object detection and walking were used by the participants in order to complete the task. Some of participants were analyzing the scenario moving from the left to the right side, detecting the two columns. After detecting them, they passed between them (see Figure 5.6 with red colour). In other alternative, the participants were analyzing the scenario looking to the left and right; when they detected the objects, they passed between them. There were significant differences between the first and the second trajectory results; in the second trajectory, the participants achieved better results both in time and in number of hits, due to the confidence on the system information and practice.

There were also individual differences in performances; some participants had little difficulty with the spatial sounds. In general, they performed very well on these trials, trying to have the best results on both tasks.

The use of acoustical representation of the environment has been selected because of the fast processing and delivery of the information, let us say in real time. Also, the blind users have well developed the auditory system. The deficit found here is that, at first, it is difficult to perceive the real object position, because the system only emits acoustical signals with 100%

accuracy, but the human brain should learn the signals. There are some errors, which amount for around 50cm. Speech technology would inform on the object position, but it would spend too much time; moreover, in the case that in the frontal plane there were more objects, the speech technology would lead to confusion in the user, whereas the acoustical system can represent several objects at the same time, without ambiguity. When the user walks towards the objects or the object is moving, or both move, the acoustical system represents the objects as getting closer or farther.

Another limitation of the system is that the environment is presented only for the frontal plane with an azimuth of 64º, since the sensor is designed for a horizontal line at the eyes level. In that case that, there were small objects in the frontal plane, the user could collide with them (if the objects do not appear on the sensor range, the system will not detect them). Because of this reason, the users need to move the head up and down. That situation is quite difficult for the blind people, because they are not used to move their head.

The error could be more significant in the real-world navigation, due to the street signals and noises. Therefore, an important future work will consist of the navigation in the real environment.

Further research is also focused on the integration of the 3D-CMOS sensor within the artificial vision system, in order to obtain the whole 3D image of the environment representing that information by short acoustical signals. The purpose of that study is the improvement of the environmental data acquisition and its presentation to the users. This will allow the user to represent the objects they need, the most dangerous, the moving objects or the free-paths, to connect both systems or to connect only the one he considers, to adjust the sound volume, etc.

Also it can be useful to investigate the tactile interface instead of the acoustical signals or both altogether; this might help people with auditory problems.

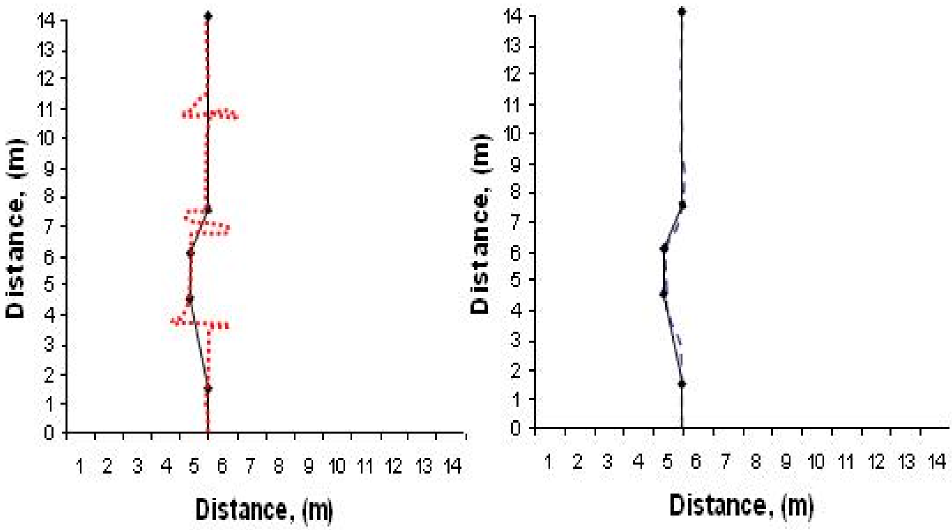


Figure 5.6 Trial trajectories for two situations. The red trajectory is for the case in which the participant was looking and analyzing the scenario moving from the left to the right. On the other hand, the case in which the participant was analyzing the situation moving the head to the left and right is represented in blue ink. The optimum trajectory is represented with black

5.2.3 Conclusions

The results presented in this section demonstrate the usefulness of the perceptual cues for navigation purposes. Also, it has been proven that the auditory spatial cues show a high effectiveness in detecting spatial objects and orientation. The sensor used, gives real time information which was represented by acoustical signals. The research provides significant results, ratifying that navigation with acoustical signals may have good performances for the blind users. Data show the direct influence of the trainings on the navigation improvement for visual impaired people. The participants showed great results on time and number of hits; a minimum time of 0,41min was needed to complete the whole trajectory for laboratory trials, whereas the number of hits decreased up to 0 for the second trajectory for almost all participants. The performances show that, after some trials, the users were able to achieve a total control of the system and to navigate safely.

5.3 Object detection and navigation tests II

This section describes the tests carried out for object detection with the cognitive aid system for blind people based on acoustical signals. Spatial signals were used to generate the cognitive task when the participants (blind users) were walking along a specific labyrinth. The information provided by the system indicated in real time the precise location of the obstacles both in distance and azimuth, by means of spatial signals. Thanks to the sensor precision and characteristics, such as those of the acoustical signal, significant results were obtained. The experiment consisted of two different parts; first, the Basic Learning Protocol, during which the participants became familiarized with the spatial signal, sound externalization and object recognition and, second, the navigation along the labyrinth. The participants showed great results both in time and in number of hits; a minimum time of 0,41 min was achieved to complete the whole trajectory for the laboratory tests, whereas the minimum number of hits decreased to 0 for the second trajectory.

The results of the experiments based on acoustical signals and time of flight laser show that the system was able to control the localization of the real obstacles both in distance as in azimuth. The results have direct implication on human everyday navigation and guidance for visual impaired people.

Besides the environment data acquisition system, an important factor for the blind user is the representation of the acquired information; this is, how to inform the blind user that the system detected an object in the front of his view.

Nowadays, several methods such as tactile, speech or acoustically sounds are being developed.

Participants

Twenty blind users with an age range between 26 and 69 years and different gender, from Germany (Group A) and Italy (Group B) participated on these experiments. All participants had normal distance estimation and good hearing abilities. They demonstrate a correct perception of the sounds through headphones. Since the experiments took place in two different countries, two identical systems were developed, so that one of these systems was used in each group. Each group each was based on ten people.

- Participants from Germany (Group A) had the following characteristics:

— Age range: between 26 and 69 years
— Sex:
Male: 6
Female: 4

— Professional Status:
Working: 2
Unemployed: 4
Retired: 4

— Type of blindness:
Congenital: 2
Adventitious: 8

— Onset of blindness:
Ages between 0 and 5 = 4
Ages between 6 and 18 = 1
Ages between 19 and 39 = 5

— Distance Estimation Ability:
Good: 6
Poor: 4

All participants reported to have good hearing. All were experienced cane users. One of the participants was a cane user who had a guide dog.

It must be taken into consideration that, in Germany, legal blindness refers to central visual acuity of 1/50 in the best eye with the best possible correction, as measured on a vision chart. Practically, this means that a blind person is able to perceive an object at a distance of 2 cm, which a sighted person will be able to perceive from a distance of 1 meter.

- Participants from Italy (Group B) had the following characteristics:

— Age average: 38,8 years
— 5 men and 5 women
— 6 people with good mobility skills and 4 with less skilled
— 2 guide dog users
— Different pathologies (glaucoma, retinis, etc)

Devices

For the navigation task, a 3D-CMOS Time of Flight sensor was mounted into a pair of glasses, and an acoustical interface was also integrated into the system. The purpose of these components is to analyse the scene in front of the user and, through an acoustical system, to inform the user about the obstacles in his direction of view. After the user had the system connected, if there were any obstacles in front of his view, the system informed on the position of the objects by means of the headphones. This position was fixed in distance, azimuth and elevation.

It is desirable for the visual information input unit to be small and lightweight since these devices will be mounted on the user head. The sensor system with all the optical components, analogue and digital electronics and laser is assembled into a pair of glasses, as shown in Figure 5.7.

The maximum distance reached by the sensor is 5 m at 64º in azimuth. The patented measurement principle is based on ToF measurement of pulse-modulated laser light using a high-speed photosensitive CMOS sensor and infrared laser pulse illumination. The analogue signals of several laser pulses are averaged on a chip in order to reduce the required laser power and also to increase measurement accuracy. A fully solid state micro system is embedded on FPGA. The advantage of using these sensors is to provide an exact distance both in a horizontal and in a frontal plane. In addition, they reduce the necessary processing time for the computation. The information from the CMOS sensor is used in the audio representation module when decisions are taken for generating the appropriate sound map.

A computer program written using Visual Basic combines the artificial vision and acoustical module written on C++, with the FPGA for the sensor into a unique platform.

The angular range of the sensor is 64º. The distance unit used is centimetre. The object coordinates are given by the artificial vision algorithm to the acoustical interface in pixels.

The distance range for the acoustical module covers from 0,5m to 5m, an azimuth of 64º, and 64 sounding pixels per image at 2 frames per second. The sounds are emitted with an inter-click interval of 8ms. A delta sound (click) of 2048 samples and sampling rate of 44.100 Hz has been used. In order to obtain the spatial sounds, a Head-Related Transfer Function filter has been used for each 1º in azimuth (for 32º left and 32º right side of the user) at each 1cm in distance. The HRTFs were simulated as precisely as possible with a KEMAR manikin. The acoustical module has been provided with an 8-

level volume control, in order to adjust the sound according to the convenience of each participant.

The sensory modules provide two main types of data on the user frontal scene: first – the location, direction of the objects, and second – the set of coordinates in which a horizontal imaginary plane, located at the eyes level, cuts the surface of the existing object. The acoustical module is able to synthesize on real time the set of sounds to be delivered, by means of a convolution operation between every spatial filter provided by the sensor system.

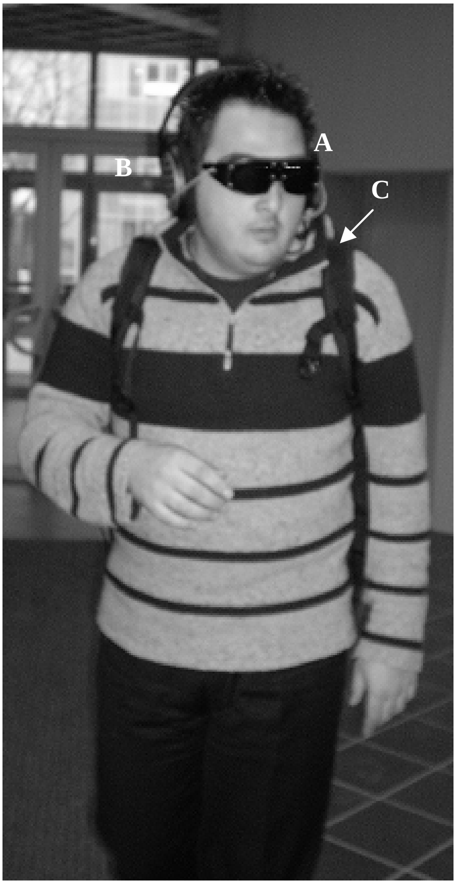


Figure 5.7 Devices for navigation task: A) 3D-CMOS laser implemented on a pair of glasses, B) Headphones, C) Backpack with the electronics

The acoustical module presents audio information to the user, representing a limited area of the subject frontal scene. This area consists of a plane, horizontal to the user head, located at the height of the ears.

High quality headphones SONY MDR-EX75SL were used to deliver the acoustical information about object position (See Figure 5.8). A diaphragm of 9mm with high sensibility helps to perceive the sounds at a more expanded scale of 103dB from 5kHz up to 23kHz. For the user comfort, an ink pad is designed with an angled structure made of silicone. The model has been selected for not interfering with environmental sounds such as cars, traffic lights, people speaking, etc. These models of headphones are intern headphones with high performances, fitted to listeners which do not want to loose sound quality.

Figure 5.8 Headphones model used in the experiment

Measured Variables

Due to the blind community requirements, four variables were measured: the Absolute Walking Time (AWT), the Walking Time (WT), the Number of Hits (NH) and Number of Corrections (NC).

The Absolute Walking Time – refers to the amount of time needed to complete the route when using the long cane.

The Walking Time – is the amount of time needed by the user in order to complete the whole route when using the system.

The Number of Hits – indicates the number of hits made by the user during the route.

The Number of Corrections - During the experiment, the users got often lost. When they were unable to find the right way, the mobility instructors intervened and repositioned them.

Procedure

The tests were designed in two different phases: familiarization with the sounds and system (phase 1; Basic Learning Protocol (BLP)), and navigation with the acoustical system (phase 2) described in the next subsections. Different tests were needed in order to complete both two different phases. Each phase is described next.

5.3.1 Phase 1: Basic Learning Protocol (BLP)

5.3.1.1 Method

Objectives

– To know whether a blind person can or can not walk from one place to another, avoiding a set of obstacles without using the touch as a reference, but using the auditory information from the device, both without and with a previous short period of training.

– To know how well can this task be completed, in terms of speed and precision (number of hits per trajectory), both without and with a previous short period of training. This is, to measure the task performance.

– To know both whether there is any improvement after a short period of training and how large such an improvement is.

For the initial practice, in order to make participants familiar with the sounds and to learn how the objects in the environment are acoustically represented, a Basic Learning Protocol (BLP) was carried out. During this exercise, the participants were realizing and understanding that the sounds received through the headphones come from the environment, representing an object. They learned and improved the sense of externalization. In this way, the brain was trained to translate the environmental objects into acoustical sounds, creating a virtual representation of the real environment.

For the Basic Learning Protocol, the participants were supposed to develop seven different exercises (Figure 7.3):

Exercise 1: a single column detection
Exercise 2: identification of a pair of columns
Exercise 3: to pass through the two columns (open door)
Exercise 4: identify two columns together
Exercise 5: to identify a wall
Exercise 6: a column in front of a wall
Exercise 7: to move towards the objects and backwards again

The Exercise 4 of the experiment is quite similar to the Exercise 2. In this exercise, the columns are the same columns used for the Experiment 2. However, in Exercise 4, the idea is to localize an object with a volume bigger that the user.

Thus, the participants were creating a map, in which each participant attributes to the obstacle location the corresponding acoustical sound, both in distance and in azimuth. Also, they were able to analyze and compare the listened sounds, to touch the object and, at the same time, to listen through the system the sounds representing that object. In order to protect the participants from any damage, the objects were made of cardboard (See Figure 5.9).

The experiments started with the user looking towards a free space, in order to hear no sound coming from the system. When facing the experimental scene, the participant heard a multitude of sounds through the headphones, always that objects appear in the front of his view. Depending on the objects number and location, a corresponding number of groups of sounds were delivered with different intensity. Thanks to the sensor characteristics, which sent laser flows at each 1º in azimuth, and to the acoustical module, which emitted spatial sounds for each 1º in azimuth and 1cm in distance, it was possible to perceive the object width as a line of sounds with the same thickness. Therefore, the participant obtained exact acoustical information both in distance and in azimuth for the object location.

The tests were developed in halls of $258m^2$ at the Schools for Blind People located in Berlin and Italy. The tests were supervised by the responsible for the (DBSV), Hans Kaltwasser (group A), and by the responsible for IFC in Italy Giovanni Ciaffonni (group B).

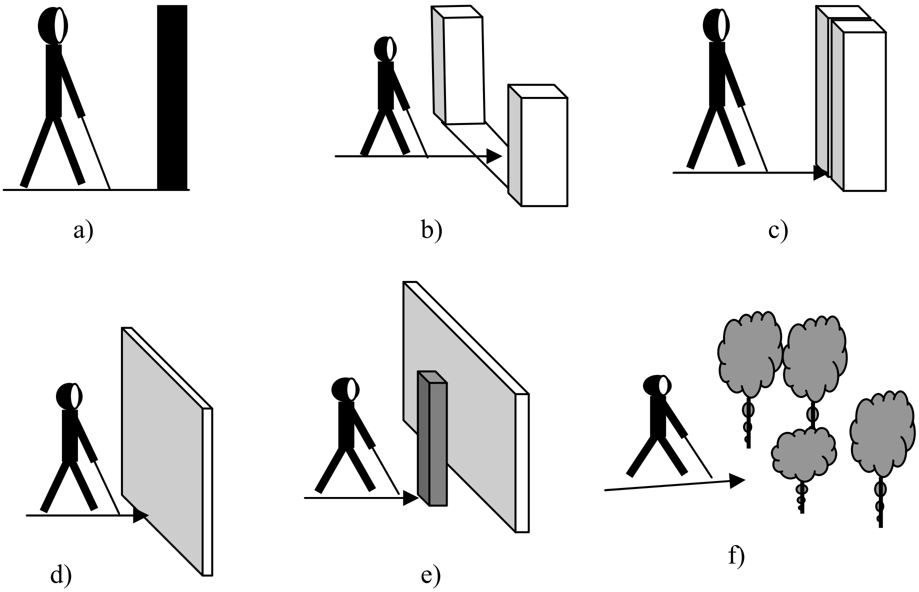


Figure 5.9 Experimental scenarios: a) Single column detection; b) Two columns detection and pass through them; c) A wide wall; d) A column detection in front of a wide wall and e) Outdoor experiment

After performing each practice with the Basic Learning Protocol, the research assistant and blind trainer were giving suggestions and answering the participant doubts.

5.3.1.2 Basic Learning Protocol results

During the BLP tests, the time for sound externalization, and object detection were registered. The results are illustrated in the Table 5.1, Table 5.2 and Table 5.3; they are also shown in the Figures 5.4, and Figure 5.5

In the Figure 5.10, the seven bars shown in each row (each row corresponds to a subject) represent the seven different exercises, as labeled. For each exercise, each blue bar shows the registered time required by each subject to complete the exercise.

Performances corresponding to the group A will be discussed firstly. When performing the <u>first exercise</u>, in which it was required to detect one single column at a distance of 2,50 meters, it can be seen that overall scores were quite weak. The average time registered was 1,35 minutes. The minimum time registered by the subjects was 0,25 minutes and the maximum of 2,17 minutes. These high required times were achieved due to the ignorance on the system working process. The subjects did not know how the sounds are, what they represent. Since the maximum registered time of 2,17 minutes for the exercise 1 corresponded to one out of 20 subjects, it was decided to calculate the population distribution for the experiment. For this purpose, a segmentation into three different groups was considered, according to the required time for completing the exercise: *good time* which includes the times between 0 and 1 minute; the *normal* time, between 1 minute and 2 minutes and, finally, the *bad time,* between 2minutes and 3minutes. 30% of subjects were included in the first group, 50% were included in the second group and only 30% in the third group. According to this classification, we can conclude that most of subjects spent between 1 and 2minutes, which means an acceptable result for the first exercise, when the subjects experiment for the first time the system.

If an additional column is added, placing it a certain distance relative to the first, as if they formed a door, we have the <u>second exercise</u>; the subjects found this exercise complex but, at the same time, simple because they understood better how the system was working. The average time registered for the second exercise was 22,63 minutes. There were 3 subjects that needed less than one minute for perceiving the open space between the two columns; the minimum registered time was 0,45 minutes. 10% of subjects required between one and two minutes, whereas 60% of the subjects requested more than two minutes to complete this exercise.

The <u>third exercise</u> is similar to the exercise two. In this exercise, the subject must detect the door and pass through it. Slight improvements were registered. The average time was 4,05 minutes; the minimum time was 2,05 minutes and the maximum 6,40 minutes. In this exercise, it could be said that the users achieved very good results, since the exercise was much more complex and the users required a certain time for walking 2,5 meters. 20% of subjects needed a time about two minutes, 50% between two and five minutes and 30% of the subjects spent between five and six minutes for detecting the object and walking to it.

In the <u>fourth exercise</u>, the subjects must detect one big object. For this purpose, the two columns were placed together. In this exercise, slight improvements were registered. The average time needed was 2,6 minutes; the minimum time was 0,34 minutes and the maximum 4,38 minutes. 10% of subjects needed a time between 0 and 1minute, 30% between 1 minute and 2minutes and 60% needed more than two minutes to detect the object.

In the <u>fifth exercise</u>, the subjects should identify a wall in front of them. All subjects showed good improvements; the average time was 3,53 minutes; the minimum time was 1,32 minutes and the maximum time was 4,51minutes. No one needed less that 1 minute to complete the exercise, 10% spent less that 2minutes and 90% requested more that 2 minutes for the whole exercise.

For the sixth exercise, which consisted of the detection of an object in front of a wall, the average registered time was 6,32 minutes (highest time registered for all exercises), the minimum time was 4,01minutes and the maximum was 8,15 minutes.

Finally, for the seventh exercise, the subjects must walk around the scenario. The minimum registered time was 3,15 minutes and the maximum 6,14 minutes; the average time was 4,52 minutes.

In Table 1 can be also seen the summary and average time, for each subject, for all seven exercises. Comparing the times needed by the subjects, we can observe that, in general, there are no significant differences between them. In some cases, some of the subjects perceived better than others.

Let us see the evolution of one of the subjects, selected randomly, without any special criterion. Let us analyze, for instance, subject TP7. In the first run, that subject registered for the first exercise 0,25 minutes, for the second run – 0,37 minutes, and for the third run 0,27 minutes. This means that this particular user required the minimum time for the first run (it should be said that the user did not know a priori the path to be followed); the average time required by subject TP7 for the first exercise was around 0,31 minutes. In the second exercise, he registered 0,45 minutes in the first run, 1,43 minutes in the second run and 0,48 minutes in the third run. From this exercise we also are unable to say if the subject improved his time. The improvements start with the fifth exercise where we can see that in the first run the registered time is 3,55 minutes, the second run improved with 1,6 minutes and the third run improved with almost 2 minutes. Generally, the subject TP7 recorded improvements of 2,22 minutes on the third run.

Let us analyze the results for the subject TP4. The time results of the first exercise do not indicate great improvements, in the first run the subject registered 1,17 minutes, in the second run he needed 0,41 minutes and in the third run he needed 0,48 minutes. In the exercise two, the subject also had some improvements during the trials; in the first run he required 3,35 minutes, in the second run he improved its time - 0,29 minutes- and the third run improved with 1,9 minutes in comparison with the first run. Generally, the subject TP4 registered an average improvement for all seven exercise of 2,53 minutes.

Due to the fact that the subjects carried out the runs in different days, it is normal that great differences arise between the results of the time registered in the three runs for the first exercises; so, they are not significant. When there were several days between the runs, the subjects forgotten how the system was working; they required every time to be instructed about the exercise. After one or two exercises the subjects got used to the system, thus they straightway scored better results. Therefore, we can mention that if the user will have the system at home where he is able to test it during great period of the day, during several weeks or months they will be able to travel safety and confidently.

An analysis of the results of the group B is presented below. For some exercises, the group B registered better results in time than the group A. Let us see the results for each exercise separately. In the exercise one, great results were registered. The minimum time registered was 0,50minutes, i.e., 0,25minutes more that the group A, the maximum time is 3,30minutes, the average time was 1,57minutes. 20% of subjects required less that one minute for the whole exercise, 30% required between one and two minutes and 50% required from two minutes up to 3,3minutes.

In the second exercise, the average time was 2,04minutes, where the minimum time was 0,25minutes and the maximum required to complete the exercise was 5minutes. 20% of the subject spent less than one minute for the whole exercise, 30% spent between one and two minutes, 50% required more than two minutes.

For the exercise three, a great difference between the average time of the group A and B was registered. The group A average time was 4,05minutes, whereas the group B only needed 1,48 minutes (thus, a difference of 2,17minutes). The minimum time spent to complete the exercise was 0,25 minutes and the maximum was 3 minutes. 40% of the subjects spent less than one minute, one subject spent between one and two minutes and 50% spent more than two minutes for the whole exercise.

In comparison with group A the group B achieved also better results for the exercise four. The average time for the exercise was 1,38 minutes for the group B, with almost one minute less; the minimum was 0,15minutes and the maximum was 3,03minutes.

In the exercise five, the average time for the group B was 2,22 minutes better. The minimum time registered was 0,25minutes and the maximum was 4,20minutes.

In the exercise six, the difference between the average time of the group B and A are substantial (around 4 minutes). Two of the subjects spent less than one minute for completing the exercise. The minimum registered time was 0,25 minutes and the maximum was 4,20 minutes.

For the exercise seven, the group B registered an average time of 2,16minutes, 0,20 minutes for the minimum time and 5,10 minutes for the maximum time.

According to the time results of the Basic Learning Protocol experiment, we can conclude that great differences were registered both between exercises

and groups. The results show that different exercises present different level of complexity. From the registered data, after the BLP finished, we can note that the exercise number 1 “Single column detection” was easy to do.
Figure 7.5 and Table 7.2 show that the required time to complete the exercise increases with the exercise complexity. For both groups, the most difficult exercise was the exercise 6 and 7 “Column detection located in front of a wall” and “Walking around the scenario”.

The group A average time for the exercise 6 was 6,32 min, whereas the group B average time was 2,13 min., so there was a substantial difference. In general, the total time spent to complete all seven exercises of the BLP by the group A was 240 minutes and for the group B was 138,45 min.

The great differences between the results of both groups are due to the experience with Electronic Travel Aids. The group B was based on blind people who had experience with different ETA systems, whereas the subjects from the group A did not have any experience. Also is should be remarked that most of the subjects of the group B lost the vision due to accidental circumstances; most of them had seen during his life how the things look like and, therefore, they had knowledge on the environmental world.

In order to carry out a comparison on the evolution of the perception and due to the high time registered during the first trial, the group A completed the BLP experiment three times. Great improvements were achieved. The improvements of the average time at the second run increased with 53 sec and the third run increased with 1,54 sec in comparison with the first run.

Figure 5.11 and Figure 5.12 show graphically the results of the first, second and third run carried out by the group A.

In the Figure 5.11, seven sets of bars in each panel represent the seven different exercises as labeled. For each exercise, the bars show the registered time required by each subject to complete the exercise. The pointed bars represent the registered time for the first run, the horizontal filled bars represent the second run results and the white/black filled bars represent the results from the third run.

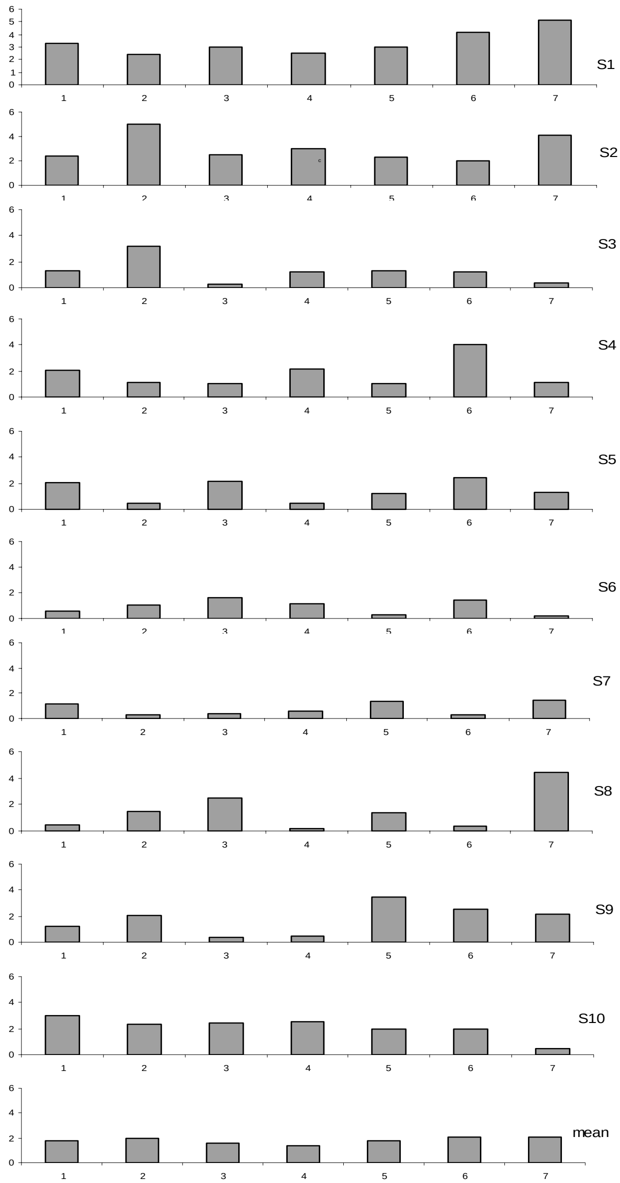
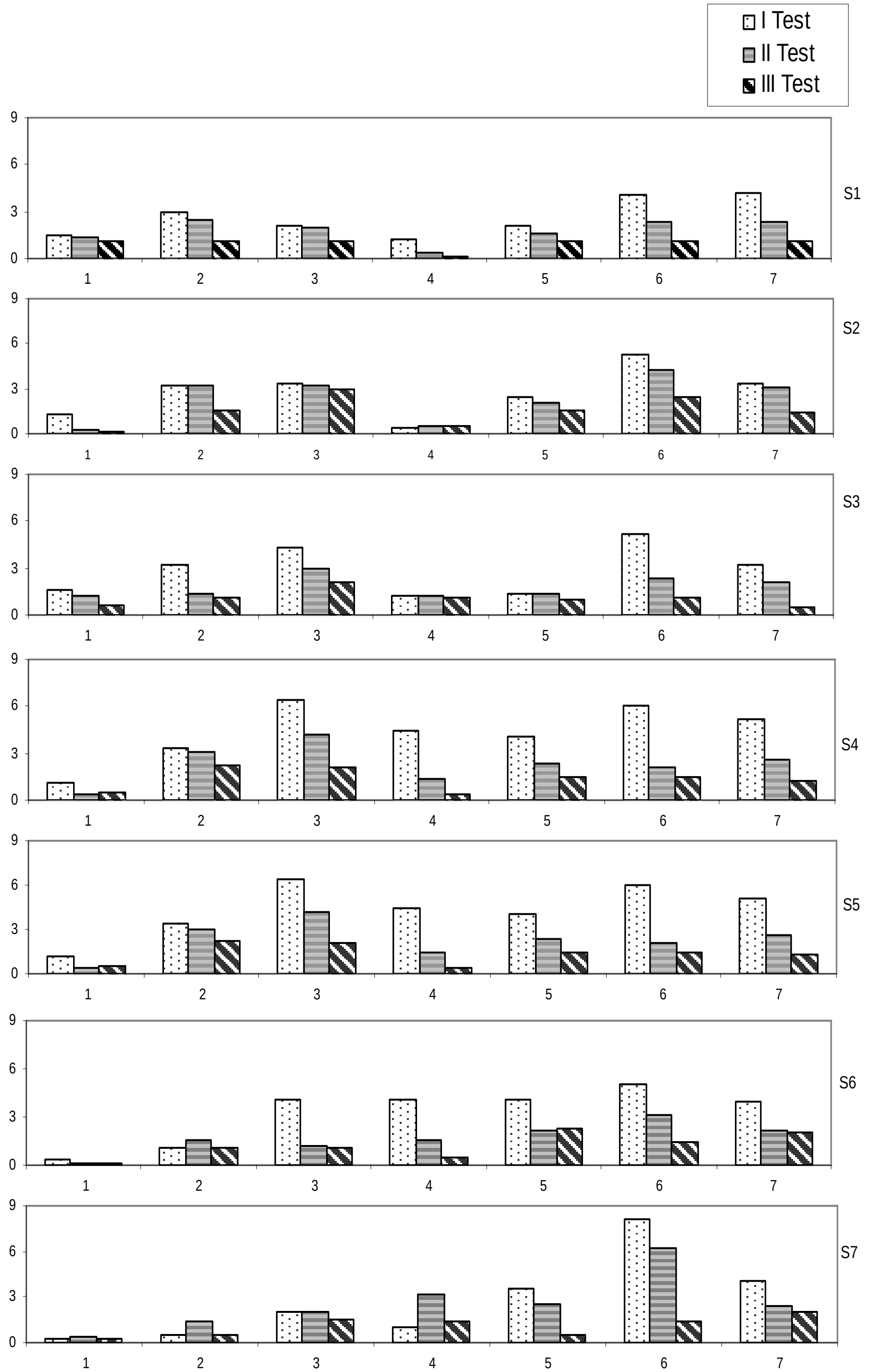


Figure 5.10 Time for object detection for the seven exercises of the group A. The horizontal axes represent the seven exercises, whereas the vertical axes show the required times in minutes. The seven exercises data from ten subjects and the average times are presented. In each panel, the results are shown for one run.


I Test
II Test
III Test
S1
S2
S3
S4
S5
S6
S7
9
6
3
0
1
2
3
4
5
6
7

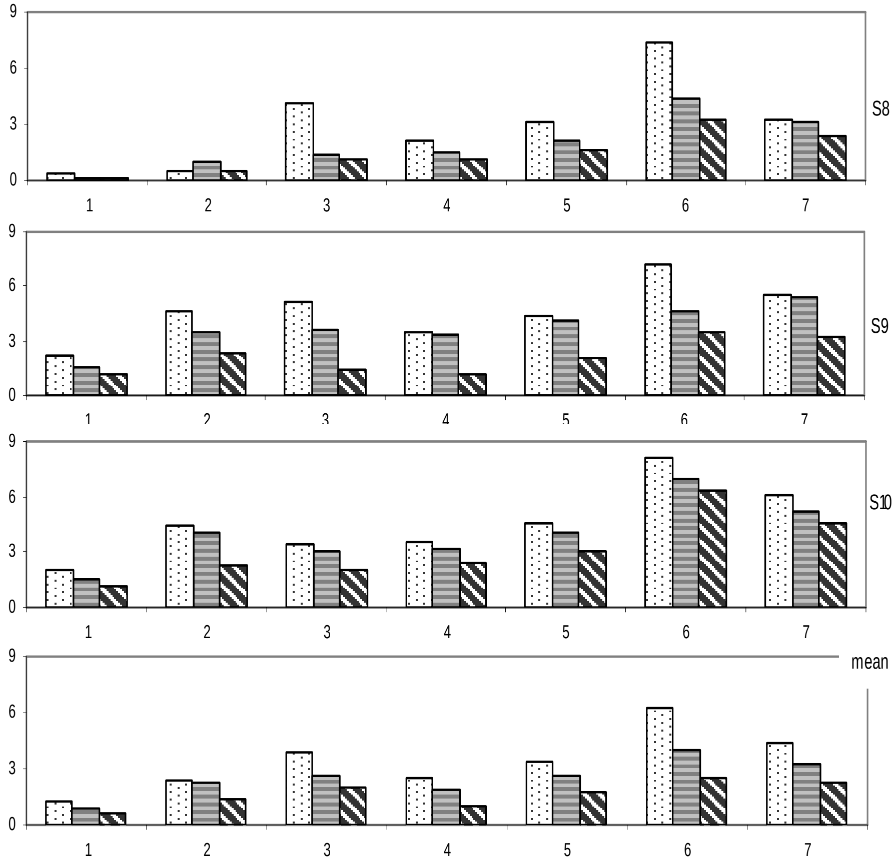


Figure 5.11 Requested time for completing seven exercises of three runs. The seven exercises data from ten subjects and the average time are labelled. In each panel, the results are shown for first run (in pointed bars), for the second run (with horizontal bars) and finally the third run (with white and black angled bars). The best time results correspond to the minimum possible values. The horizontal axes represent the seven exercises, whereas the vertical axes show the time in minutes.

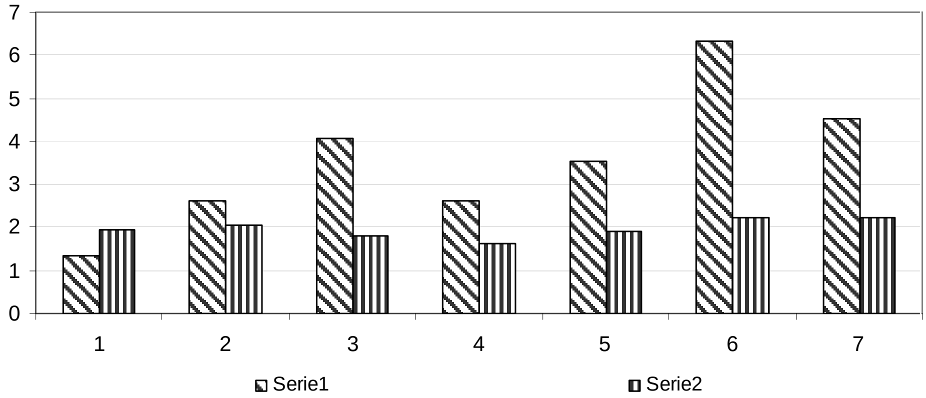


Figure 5.12 The average time of the first run for all seven exercises from the groups A and B

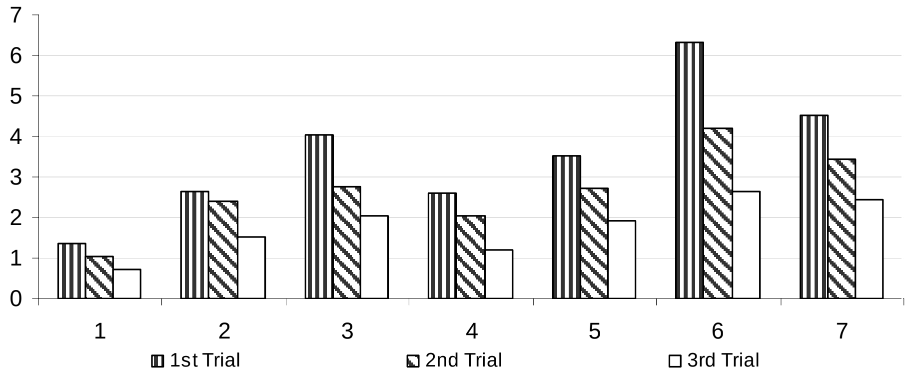


Figure 5.13 The average time for the seven exercises, for three runs carried out by the group A from DBSV

In the Figure 5.13, the results of all three runs are presented; the differences between all seven exercises can be seen. There are important differences in the time results between trials. From the results, we can see that with each trial, the running time spectacularly decreases. If for the first run of the test the users required an average time of 3,29min, for the second run the time results improved in 1,316min. The difference between the average time of the second run and the third one is of 1,65min.

Table 5.1 Results from the three trials of the Group A

First trial

Ex	TP 1	TP 2	TP 3	TP 4	TP 5	TP 6	TP 7	TP 8	TP 9	TP 10	∑	*average*
1	1,48	1,23	1,58	1,17	0,35	1,27	0,25	0,38	2,17	2,01	**13,49**	**1,35**
2	3,01	3,18	3,16	3,35	1,12	0,48	0,45	0,52	4,57	4,48	**26,30**	**2,63**
3	2,14	3,40	4,29	6,40	4,07	4,23	2,05	4,13	5,18	3,41	**40,50**	**4,05**
4	1,22	0,34	1,21	4,38	4,04	3,02	1,06	2,14	3,46	3,56	**26,03**	**2,60**
5	2,14	2,42	1,32	4,01	4,14	4,06	3,55	3,13	4,37	4,51	**35,25**	**3,53**
6	4,01	5,28	5,14	6,01	5,02	6,15	8,05	7,38	7,16	8,15	**63,15**	**6,32**
7	4,24	3,30	3,15	5,14	4,02	5,04	4,11	3,28	5,58	6,14	**45,20**	**4,52**
∑	**19,04**	**20,35**	**21,05**	**31,26**	**23,16**	**23,16**	**25,05**	**20,32**	**22,16**	**33,46**	***240,21***	
Average	**2,43**	**2,56**	**3,00**	**4,29**	**3,19**	**3,19**	**3,35**	**2,56**	**3,10**	**4,49**	**3,34**	

Second trial

Ex	TP 1	TP 2	TP 3	TP 4	TP 5	TP 6	TP 7	TP 8	TP 9	TP 10	∑	*average*
1	1,32	0,32	1,25	0,41	0,16	1,48	0,37	0,18	1,56	1,48	**10,53**	***1,05***
2	2,48	3,16	1,36	3,06	1,58	1,02	1,43	1,02	3,48	4,01	**24,20**	***2,42***
3	2,01	3,19	3,01	4,21	1,17	3,15	2,01	1,32	3,56	3,07	**27,51**	***2,75***
4	0,39	0,50	1,27	1,38	1,52	2,06	3,18	1,48	3,37	3,16	**20,31**	***2,03***
5	1,56	2,02	1,31	2,38	2,18	3,46	2,49	2,07	4,16	4,07	**27,20**	***2,72***
3	2,32	4,20	2,35	2,11	3,17	4,16	6,17	4,34	4,57	7,01	**42,00**	***4,20***
7	2,37	3,07	2,06	2,56	2,13	4,58	2,46	3,07	5,42	5,18	**34,50**	***3,45***
∑	**14,05**	**17,22**	**13,41**	**17,07**	**13,11**	**21,01**	**19,31**	**14,19**	**28,12**	**28,38**	**187,45**	
Average	**2,01**	**2,29**	**1,57**	**2,30**	**1,53**	**3,00**	**2,47**	**2,40**	**4,10**	**4,50**	**3,70**	

Third trial

Ex	TP 1	TP 2	TP 3	TP 4	TP 5	TP 6	TP 7	TP 8	TP 9	TP 10	∑	*average*
1	1,05	0,18	0,63	0,48	0,14	0,56	0,27	0,16	1,12	1,13	**7,33**	***0,73***
2	1,17	1,56	1,11	2,26	1,10	0,52	0,48	0,56	2,28	2,34	**15,39**	***1,54***
3	1,10	3,00	2,05	2,12	1,13	4,26	1,46	1,12	1,37	2,01	**20,42**	***2,04***
4	0,16	0,46	1,06	0,43	0,45	1,39	1,42	1,16	1,17	2,47	**12,17**	***1,22***
5	1,12	1,52	1,02	1,43	2,24	2,58	0,56	1,59	2,02	3,07	**19,15**	***1,92***
6	1,05	2,42	1,16	1,42	1,46	2,48	1,39	3,29	3,46	6,28	**26,39**	***2,64***
7	1,06	1,45	0,54	1,28	2,01	4,16	2,01	2,36	3,24	4,56	**24,27**	***2,43***
∑	**7,11**	**12,19**	**8,37**	**11,20**	**9,34**	**17,51**	**9,19**	**11,45**	**15,46**	**23,60**	**126,34**	
Average	**1,10**	**1,45**	**1,17**	**1,36**	**1,15**	**2,16**	**1,13**	**1,32**	**2,40**	**3,10**	**2,60**	

Regarding Figure 5.13, we can conclude that the average time was improved in 1,49min. The improvements of the time are due to the fact that the users got used to the device. They learned how the system worked and to which things they should pay especial attention. It is necessary to remark, that the users have repeated the exercise during several days. We could imagine what should occur if the blind participants tested the system during several hours per day during one week. If the results of the three trials registered during different days are remarkable, then for the everyday usage of the device the results would improve exponentially.

After two months of trainings and having the results from the subjects, the visual impaired trainers and research specialists evaluated the travel performances and created the basis for the development of the improved acoustical system.

Table 5.2 Results from one trial of the Group B

First trial

Ex	TP 1	TP 2	TP 3	TP 4	TP 5	TP 6	TP 7	TP 8	TP 9	TP 10	Average
1	3,30	2,45	1,27	2,10	2,10	0,59	1,20	0,50	1,20	3,00	**1,93**
2	2,40	5,00	3,15	1,10	0,45	1,05	0,25	1,45	2,05	2,30	**2,04**
3	3,00	2,50	0,25	1,00	2,12	1,58	0,35	2,45	0,40	2,40	**1,81**
4	2,50	3,03	1,20	2,20	0,50	1,15	0,55	0,15	0,45	2,50	**1,62**
5	3,00	2,34	1,35	1,07	1,25	0,25	1,38	1,40	3,45	2,00	**1,91**
6	4,20	2,05	1,25	4,06	2,46	1,40	0,25	0,35	2,52	2,00	**2,21**
7	5,10	4,11	0,40	1,10	1,32	0,20	1,45	4,45	2,20	0,50	**2,24**
∑	**23,50**	**21,48**	**8,87**	**12,63**	**10,20**	**6,22**	**5,43**	**10,75**	**12,27**	**14,70**	***23,02***
X	**3,36**	**3,07**	**1,27**	**1,80**	**1,46**	**0,89**	**0,78**	**1,54**	**1,75**	**2,10**	

The improved version of the system was tested by four subjects; three of them had tested the previous system. The results of this test are presented in the Table 5.3. It is difficult to reach a conclusion regarding travel performances. The users commented that they perceived great improvements on the system. They perceived the sound more pleasant and more helpful. Regarding object detection, there was a higher precision.

Table 5.3 Results from one test for the four subjects with the improved version of the system

Ex	TP 1	TP 2	TP 3	TP 4
1	1,26	0,18	0,25	0,26
2	4,57	0,32	1,20	1,80
3	0,34	0,58	1,09	0,70
4	1,23	0,46	0,31	1,30
5	0,35	1,19	1,45	1,20
3	2,14	0,41	1,46	2,00
7	0,46	0,39	1,35	2,35
∑	**10,35**	**3,53**	**7,11**	**9,61**

5.3.1.3 Conclusions

The results presented in this section demonstrate the usefulness of the perceptual cues for navigation purposes. Also, it is proved that the auditory spatial cues show a high effectiveness in detecting spatial objects and orientation. The sensor used gives real time information which was represented by acoustical signals. The research provides convincing results, ratifying that navigation with acoustical signals may have good performances for the blind users. Data show the direct influence of the trainings on the navigation improvement for visual impaired people. The participants showed great results on time. All test subjects reported that the externalisation effect took on within a few seconds. They were able to correctly identify the dimensions of objects easily and in a relatively short span of time. Test subjects had no discernible difficulties neither in gauging the width and height of the objects or a gap between two flat objects simulating an open door, nor in walking through the gap, turning around and walking back to the point from where they had started. There were differences, however, regarding individual performances during the different exercises of the BLP. The total amount of time varied considerably. That performances show that, after some trials, the users were able to achieve a total control of the system and to navigate safely.

In comparison with the M1 tests carried out with an older version some months earlier, it is interesting to remark that the total amount of time needed by the test subjects to complete the exercises had substantially decreased. This indicates the level of improvement achieved with the upgraded version which obviously made it easier to master the challenges of the BLP.

5.3.2 Phase 2: Laboratory Tests

5.3.2.1 Experimental method

Afterwards, when the participants got a certain experience with all seven exercises, described in the precedent section, they were supposed to carry out the labyrinth test – localization of the obstacle and navigation through the labyrinth.

The experiment used the same stimulus and devices used in the previous experiment -BLP.

The participants were guided towards a free space, in which there were no obstacles in the front of view in an area of 15m^2. After the device had been connected and prepared, the user was faced to the scenario. The participants were requested to navigate through a 14m long labyrinth, with four pairs of soft obstacles of 180cm height and 70cm of thickness. The columns were placed set up in an asymmetric order, separated horizontally by 2m and at a distance of 2,5m between the last pair of columns and the wall (Figure 5.14 and Figure 5.15). The objects were based on soft boxes made of carton.

The subjects carried out several test runs under different conditions:

- using the white cane
- using CASBliP device
- using device and white cane
- after using the white cane

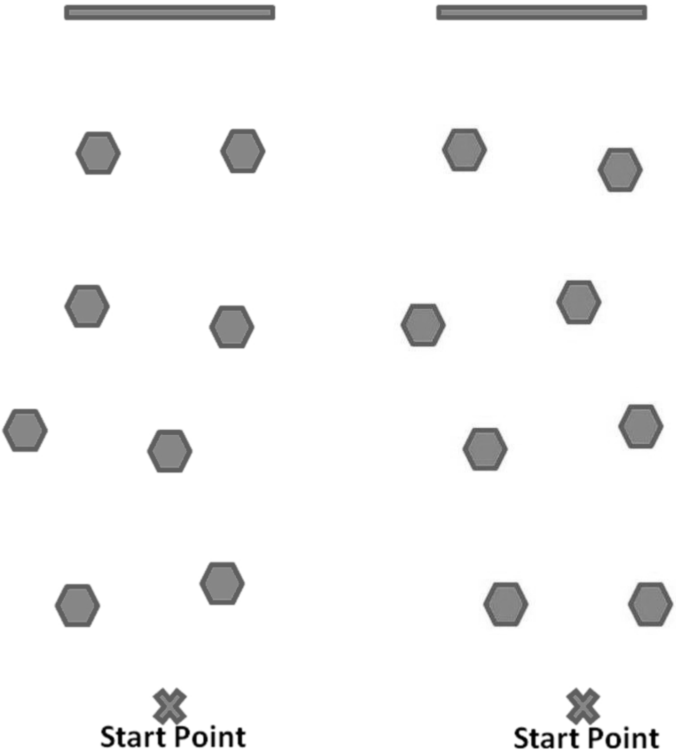


Figure 5.14 Experimental scenario

Figure 5.15 The real scenario of the experiment. The scenario was located in a hall of $258m^2$. The trajectory is created by soft-box carton columns; a square made by carton boxes placed at the hall background was used as a wall.

Two routes were prepared for the experiments. The second route represents the opposite direction to the first, i.e., the way back direction of the first route.

The whole process was recorded by video cameras from different points.

Starting the run, firstly, the user was analyzing the environment, detecting the first pair of columns; he walked among them up to the next step where it was necessary to detect the next two pair of columns passing between them etc., until completing the whole labyrinth.

When the object was detected, the participants listened spatialized sounds through the headphones, which indicated the object presence; this allowed him to avoid it and pass beside. The sound used on the navigation test was the same as for the BLP, in order not to cause confusion on the blind user.

Responses from 21 subjects for two routes were collected and analyzed, in order to evaluate the travel performances. The response was considered correct only if no obstacle was knocked and the subject reported its presence correctly. For each trajectory configuration, the obstacles were placed in a different way; performances were compared between groups. This provided information regarding the benefit provided by the subject with regards to object detection and spatial sound perception for two different trajectories.

When rating performances in time, it is important to define what constitutes the “velocity” level of performances. The subject must correctly identify the object position and its sonification used in the experiment. In the simplest view, they had a 1/8 chance of guessing the object position and avoid it. However, if one or more of objects were audible, the situation changes. For example, if one object was placed in the back of a precedent object, looking from the frontal side, the subject was detecting only one object. When the subject was moved from his position, and the two objects appeared in the area of vision of the system, then the subject perceived two objects. Moreover, when looking from an angle where the two objects are very near (up to 10cm), the subjects were not able to detect the space between them.

5.3.2.2 Results

In these experiments, the level of complexity and number of scenario configurations were progressively increasing. The data are presented for a clear inspection in Table 5.4 and Figure 5.10. Two main parameters, travel time and number of hits, were collected and analyzed. Note that after the training exercise, the participants performed the subsequent experiments without any white cane, trainer instructions or support. Beside that, the absolute walking time using the white cane, was measured after the experiment had been completed. Taking a close look at the walking time, it became obvious that the subjects had fairly great difficulties in travelling the route through the trajectory without the white cane.

In Table 5.4, AWT0 represents the *Absolute Walking Time*, the amount of time needed by the subject in order to complete the route using the white cane. AWT1 and AWT2 indicate, in each group, the time results of the first and second trajectory. NUH-1 and NUH-2 show the average number of hits for both trajectories. Beside the number of hits, in the group A, the number of corrections NOC for both trajectories is registered. The corrections were given when the participant lost the direction, so that he was not able to find the path and required help from the instructor or specialists who supervised the trial. The situations in which the participant was passing almost touching the obstacle were not considered as hits.

Individual data for five runs are shown in Figure 5.16 as well as the average data for four experimental trials. The x-axis shows the number of runs carried out in the experiment. The experiments are marked with numbers; the number 1 represents the run with the white cane, in which the subjects performed a 14m linear trajectory without any obstacles (only the group A).

The runs number 2, 3 and 4 and 5 represent the runs with the CASBliP system.

Table 5.4 Results from 20 subjects for two runs for laboratory navigation experiment

GROUP A LAB_EXP 1st & 2nd Run											*Mean*
	TP 1	**TP 2**	**TP 3**	**TP 4**	**TP 5**	**TP 6**	**TP 7**	**TP 8**	**TP 9**	**TP 10**	
AWT-0	0,28	0,29	0,35	0,26	0,25	0,40	0,28	0,32	0,47	0,42	***0,33***
AWT-1	2,45	6,58	4,21	6,45	3,18	6,01	4,15	4,32	7,31	7,02	**5,10**
AWT-11	2,23	3,06	2,17	4,48	2,14	5,27	2,25	3,29	6,58	6,54	**3,48**
NUH-1	2	4	1	3	3	5	4	3	4	3	
NOC-1	0	2	0	3	0	1	1	2	1	0	
NUH-11	2	1	0	2	2	3	2	1	2	2	
NOC-11	0	2	0	3	0	1	1	2	1	0	
GROUP B LAB_EXP 1st & 2nd Run											***Mean***
	TP11	**TP12**	**TP13**	**TP14**	**TP15**	**TP16**	**TP17**	**TP18**	**TP19**	**TP20**	
AWT-1	1,14	2,12	1,54	2,15	3,4	3,1	3,1	3,5	2,2	3,51	**2,46**
AWT-2	1,21	1,28	1,09	1,55	1,56	1,46	5,45	3,25	2,4	2,25	**2,29**
AWT-11	5,54	2,25	0,45	2	3,27	1,51	3,39	3,4	4,35	3,2	**3,09**
AWT 12	1,56	0,58	0,3	1,2	1,34	1,1	2	4,35	3,45	2,2	**2,01**
NUH-1	1	0	3	1	0	0	1	1	1	0	
NUH-2	1	2	0	1	0	0	1	1	3	0	
NUH-11	1	1	0	0	0	0	1	1	1	0	
NUH - 12	0	0	2	1	0	0	0	0	1	0	

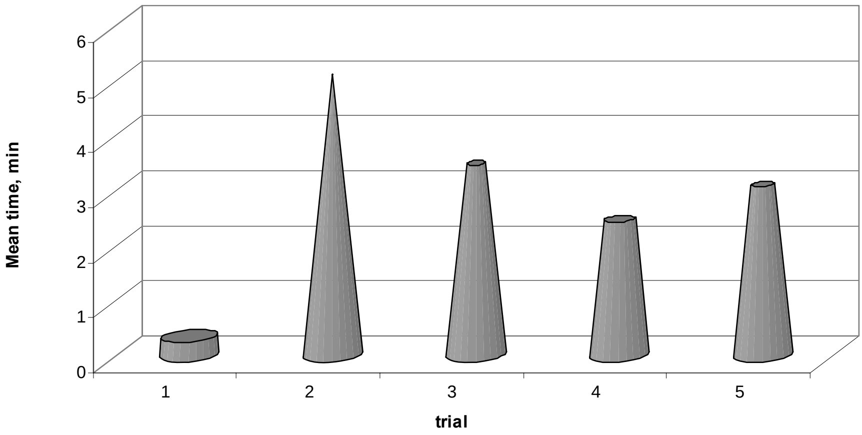


Figure 5.16 Registered walking time for four different runs of the groups A and B. In the case 1) the blind user completed the 14m labyrinth with the white cane. The 2, 3 and 4 are the runs with Acoustic Prototype.

The value of time in minutes in which the subject reported the end of the trajectory is plotted in y-axis. In Figure 5.16, it can be seen that the cone of the first experiment has a very small value. These almost perfect results are achieved due to the fact that the subjects were informed that the trajectory was a free field, where no obstacles were present in the front of view. The blind subject was walking with the white cane confident that there was no danger. Knowing this, the subject walked at the maximum speed. In the Table 5.4, it can be observed that only the group A measured the Absolute Walking Time with the white cane AWT-0.

Regarding the results from the Figure 5.16, it can be seen that the group A required a longer time in order to complete the first trajectory, almost double time than the group B. In the second trajectory, the group A registered slight improvements, whereas the group B increased the required time.

Analysing Table 5.4, it can be observed that the group B made a repetition of both runs with the aim to see if the subjects perform any improvements in time and number of hits. Small differences can be observed in the Figure 5.17, where the x-axis presents the number of runs and the y-axis represents the average time in minutes.

Note that there is a very slight improvement for the first run of the trajectory 1 and 2 and a higher improvement for the second run on the trajectory 3 and 4. These improvements are due to the subject ability of learning the system functionality and getting used to the system.

It can be seen clearly from Figure 5.17 and Table 5.4, that no improvements were noted on the case 3 regarding the case 1. It is because the user perceived the trajectories differently.

Regarding the number of hits, we can mention that with the 8 columns placed with a mutual separation of 2,5 m, the maximum number of hits was registered by the subject TP6 from the group A in the first trajectory. He hit 5 times and had one correction. The subject TP6 was one of the subjects which required longer times and had great difficulties in sound localization and object detection. He had difficulties with orientation. These results are due to the fact that the subject TP6 was total blind since born and he tested for the first time an Electronic Travel Aid System.

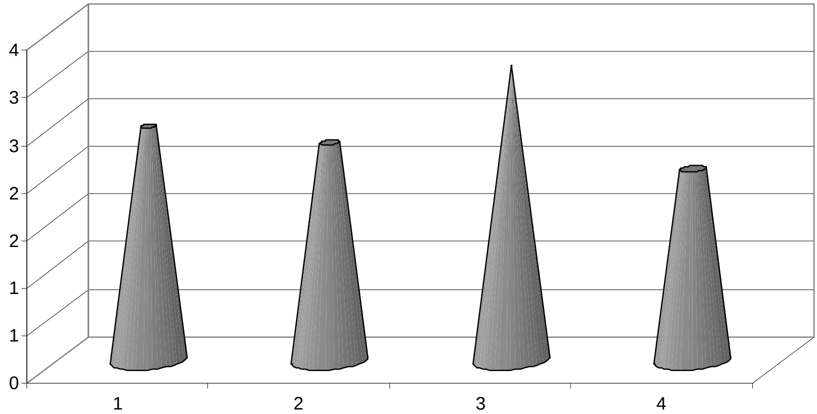


Figure 5.17 Registered walking time for two runs of the group B. In cases 1 and 3, the blind user completed the 14m labyrinth from the first time. Case 2 and 4 correspond to the repeated run with Acoustic Prototype.

Some of subjects hit due to the fact that they wanted to travel quickly; however, other subjects, such as the subject TP6, had a bad orientation or did not understand the exercise objective.

Two of the best results, both in time and in lower number of hits, were registered by the subjects TP15 and TP16 from the group B. These two subjects registered great performances on the system managing. They learned how to use and how the system was working. These subjects did not hit any columns and did not require any corrections. They were very attentive during all experiment. We notice that from the Group A best accuracy on navigation was obtained by the subject TP3. In the first run the subject made only one error, when in the second run 100% of obstacles were accurately identified and avoided, also he did not distracted from the trajectory. Good results also were noted on the Absolute Walking Time, where at the second run the results increased twice.

5.3.2.3 Conclusions

An important finding on this experiment was that object representation through spatial sounds is an easy task for the blind people. It is well known that blind people make maximum usage of the auditory system when navigate through known and unknown environment. Another finding of this experiment was that blind subjects were able to localize simultaneous sound sources and decide the location of each one. Small errors were perceived on navigation accuracy which can be explained with the training lack and attention. Comparing the results between the Preliminary tests II and the Laboratory test Phase II, it can be observed that slight improvements are registered, the maximum time registered in the Preliminary test II is 7,31min and at the Laboratory test Phase II is 5,10min. The improvements are due to the improvement of the acoustic algorithm and training. Some of the subject participated in the Preliminary test II participated in the laboratory test phase II. Regarding navigation accuracy, from the 18 possible errors, loose of walking direction or hits, the maximum errors in the Preliminary test II is 5 when in the Laboratory test II is 6.

5.3.3 Mobility Tests I

Previous experiments, Basic Learning protocol and Laboratory Test, analyzed the static object localization via acoustic signals for laboratory environment. The applied methods for object detection were evaluated through various exercises.

The purpose of the Mobility Test I was to test perceptual authenticity of the navigation, to find out how blind user were able to detect the objects and navigate through a 29m scenario in an outside environment.

In the evaluation of object detection and spatial sound localization in an open environment, there is always the problem that the sound cannot be perceived exactly as in the laboratory. That is that in the laboratory, the user is protected by the external noises as traffic noises, human, animal and bird's speech, etc.

5.3.3.1 Procedure

The test was designed as a multiple comparison task using the system as illustrated in Figure 5.18. The method used the above mentioned procedures. The listener's task was to localize the objects and to judge the differences between the perceived objects, to analyze the environment, to chose the free path and walk through the scenario. Also the perception of the objects with two different sounds was analyzed. Unideal properties of external noises influence with a slight degradation in the object detection.

In the Mobility Test I, the experiment was done by first creating an artificial scenario and then by the navigation test. The Mobility Test I was designed to correspond as closely as possible to the real environment using real and artificial objects. The test was conducted in an open and unknown environment.

Four different scenarios were used for this test. The group A, from Germany, used two external scenarios. Both scenarios taken place in the ABSV patio, the school for blind people from Berlin. The blind user should navigate through 29m long way where a variety of obstacles were placed, from the school entrance up to the school door. The test started with the blind user looking to a wide space where no object intersects the user direction of view. Because no objects appear in the system direction of view, no signals will be sent by the system to the user. Thus the user will know that in the area of 5 meters no obstacles are. Also, the user will know that when any sound will listen, then objects are in his area of view. Despite the total silence of the system the user know that this silence does not mean that the system does not work and must not to be worried.

The group B used also two different scenarios. The tests were carried out in the Institute courtyard. The scenario is illustrated in the Figure 5.18.

Both groups spent two sessions of these test (in two different days). The location of the soft obstacles was modified with exception of the real obstacles as building wall, columns and bench.

During the experiment the users were able to adjust the sound volume by their auditory necessity. They were no able to switch off the device during the experiment. During the test, the user was not allowed to ask the trainer about his position and if goes at the right way. When the user was walking in an erroneous direction and could not return to their route, then is when the trainers repositioned them.

The test assumed that the listeners had previous experience with the system and had idea of what the reproduced acoustical environments should sound like and they would be able to judge the environment.

Before the actual test, the listeners were familiarized with the task and the stimuli in previous tests, also with thee navigation task. The listeners were asked to pay special attention to a selection of attributes and describe them verbally. The attributes to be considered in the test were:

- Sound source externalization (perception of the sound source as coming from the environment and not from the headphones)
- Object detection (localization)
- Perceived distance
- Object width
- External noises (does the external noises disturb the user perception)
- Psychoacoustic practice (does the user perceive the objects instinctively, perceive its temperature or listen it, such as the blind people have the auditory very developed).
- Reflection

Generally, these attributes relation to the judged overall may differ from listener to another. The listed attributes was considered important in order to unify the evaluation. Otherwise, each listener could have concentrated on a different subset of attributes, thus neglecting others respectively.

The conductor of the experiment and the blind instructors were well aware of the imposing their own opinions on the Mobility test I through commenting on the process. Despite the explanation, the experiment conductors, a short demonstration was considered important in order to familiarize the mobility teachers with the test. Also previous exercises were carried out with the blind users in order to familiarize the subjects with the listened attributes. With these preparations, the subjects were assumed to know better where to direct their attention.

Figure 5.18 Example of the used mobility test I scenario of the group B.

After the familiarization, the listeners were allowed to freely look around, to detect objects and hear them for about 5 minutes. The free listening was followed by the full test session. The whole procedure including the familiarization session was fairly long, taking around 2 hours per listener on average. Each sample was evaluated only once by each listener. A break of ten-fifteen minutes was held after the test finish.

5.3.3.2 Participants

A total of 20 blind users completed the test. All subjects had at least some interest in and experienced with other navigation devices. All of them had normal hearing. All subjects participated in the previous experiments. Because of the difficulty of traveling the twenty blind users were organized in two groups, one in Italy composed by ten blind subjects and the another ten subjects in Germany. Each group had by one identical device.

Participants from Germany (Group A) have the next data:

In Germany legal blindness refers to central visual acuity of 1/50 in the better eye with the best possible correction, as measured on a vision chart. Practically, this means that a blind person is able to perceive an object at a distance of 2 cm which a sighted person will be able to perceive from a distance of 1 meter.

- Age range: 26 to 69 years
- Sex: male: 6 – female: 4
- Professional Status: working: 2 – unemployed: 4 – retired: 4
- Type of blindness: congenital: 2 – adventitious: 8
- Onset of blindness: ages 0 to 5 = 4, ages 6 to 18 = 1, ages 19 to 39 = 5
- Distance Estimation Ability: good: 6 – poor: 4

All test persons reported to have good hearing. All were experienced cane users. One test person was a cane user who had a guide dog.

Participants from Italy (Group B):

- Age average: 38,8 years
- 5 men and 5 women
- 6 people with good mobility skills and 4 with less skilled
- 2 guide dog users
- Different pathologies (glaucoma, retinis, etc)

5.3.3.3 Stimuli

For all tests a selection of spatialized sounds was chosen which correspond to each spatial point described in the Chapter 2. These sounds were identified to the participants at the beginning of the test. The same collection of sounds served for all sixty trials of the test. In Figure 5.19 can bee seen the graphical evaluation of one of the used in the tests sounds.

<u>The recording system</u>

- The bit depth for recording: from 16 to 24 bits per sample
- The achievable sample rate: increased to record at 96 and 192 kHz

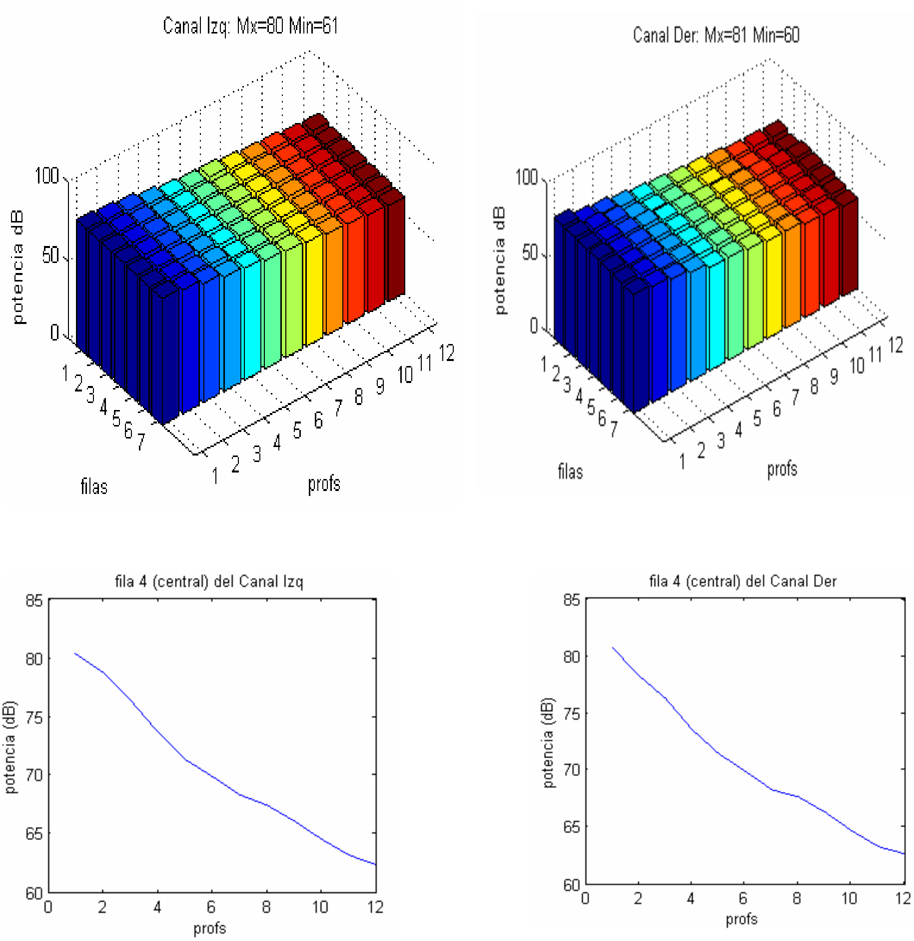


Figure 5.19 Example of graphical evaluation: Amplitude (dB) per distance (1-12) and per elevation (1-7) for one azimuth (0º, column 7) and for both channels left and right. The array of 13x7x12, at 48kHz

5.3.3.4 Mobility Test I results

Table 5.5 shows the mean time for each run under two trajectories of this experiment as well as the number of heats and number of corrections. The time was highest for the first run and for almost all subjects an increasing improvement with the second run was noted. Time performances bars are illustrated in the Figure 5.20. Abscissa represents the trials where the number 1 represent the first run of the group A, the number 2 is the repeated trajectory by the group A. The number 3 represent the men time for the first trajectory of the group B, the number 4 show the mean time of the repeated trajectory by the group B. The 5 number illustrate the mean time of second trajectory completed by the group B and finally the number 6 represent the mean time of the repeated trajectory by the group B.

Table 5.5 Results from the 20 subjects for the mobility test

MOB ILITY TEST I_EXP 1st & 2nd Run											Mean time
DBSV	TP 1	TP 2	TP 3	TP 4	TP 5	TP 6	TP 7	TP 8	TP 9	TP 10	
AWT-0	0,50	0,58	1,12	1,07	0,52	1,25	1,08	1,11	1,43	1,33	1,00
AWT-1	5,07	7,14	8,46	7,04	5,22	12,18	7,07	7,34	15,16	14,31	**9,02**
AWT-11	4	4,2	7,09	6,01	3,35	11,1	5,03	7,1	14,07	12,34	**7,31**
NUH-1	2	2	1	2	2	4	3	4	4	4	2,80
NOC-1	1	1	1	1	1	0	2	1	3	3	1,40
NUH-11	1	1	3	1	1	3	1	4	4	1	2,00
NOC-11	1	0	2	1	0	1	1	2	2	2	1,20

IFC	TP 11	TP 12	TP 13	TP 14	TP 15	TP 16	TP 17	TP 18	TP 19	TP 20	Mean time
AWT-2	8,07	9,10	2,19	18,00	6,29	8,20	8,05	6,25	8,27	9,10	**8,27**
AWT-21	6,41	8,03	2	11	4,44	6,4	7,52	5,45	8,01	6,5	**6,46**
AWT-3	6,3	7,23	5,38	4,55	8,37	7,4	9,01	4,3	8,1	5,45	**6,49**
AWT-31	5,2	5,28	4,05	7,29	5,35	7,1	8,3	3,1	7,4	4,4	**5,55**
NUH-2	0	0	1	0	0	1	1	3	1	0	0,70
NUH-21	0	0	0	0	1	0	0	1	0	0	0,20
NUH-3	1	1	0	0	1	1	0	0	1	1	0,60
NOC-31	0	1	1	2	4	0	0	0	0	2	1,00

Generally the group A required 90,19 minutes for the first run. For the second run they spent 75,09 minutes. The group B registered better results with respect to the total and mean time. For the first run the group B spent 84,32 minutes, for the run 2 – 67,37 minutes, for the 3rd run 68,09 minutes and for the 4th 59,07 minutes. Regarding the Table 7.5 and Figure 5.20 we can confirm that a slight improvement was detected for on the repeated run. Also we can observe that all subjects showed a very similar pattern on the first run, with exception of four subjects which had great difficulties on orientation.

Three of them beside to required more time to navigate through the trajectory, they was less accurately in the navigation. These subjects TP6, TP9 and TP10 performed difficulties in object detection making more hits and confusions and needed more corrections. In what regarding the subject TP18 which spent the highest time for the experiment, he registered a great accuracy in the navigation, he did not heat any object and also did none needed any corrections. The subject TP18 preferred to spend more time in navigate and analyze the surrounding scene and to make a qualitative exercise. For example, the subject TP18, showed a dynamic increase on time when repeated the run. For the second run, the subject felt more confident in the system functionality and performed the run almost 3 times faster with any hits and corrections.

Figure 5.21 show the average number of heats for all subjects for the first and second run. The 1 and 2 bars represent the average value of the group A where the bare 1 shows the average values for the first run and the bare 2 represent the average value of the second run. The bare 3 and 4 shows the average data for the group B respectively, where the bare 3 plot the average value of the first run and the bare 4 the results of the second run.

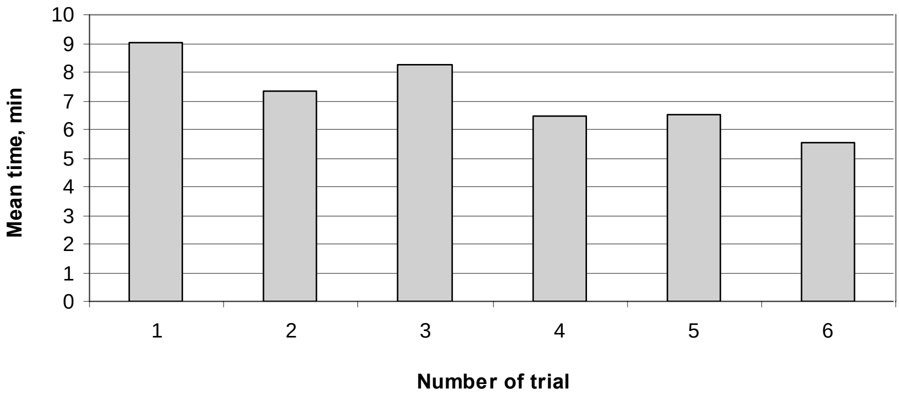


Figure 5.20 Time performances for object detection. Abscissa represents the trials where the number 1 represent the first run of the group A, the number 2 is the repeated trajectory by the group A. The number 3 represent the men time for the first trajectory of the group B, the number 4 show the mean time of the repeated trajectory by the group B. The 5 number illustrate the mean time of second trajectory completed by the group B and finally the number 6 represent the mean time of the repeated trajectory by the group b.

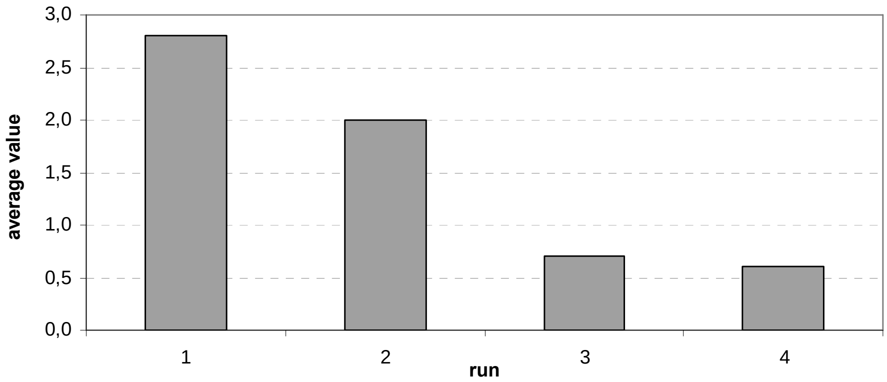


Figure 5.21 Average number of heats for all subjects for the first and second run. The 1 and 2 bars represent the average value of the group A where the bare 1 shows the average values for the first run and the bare 2 represent the average value of the second run. The bare 3 and 4 shows the average data for the group B respectively, where the bare 3 plot the average value of the first run and the bare 4 the results of the second run.

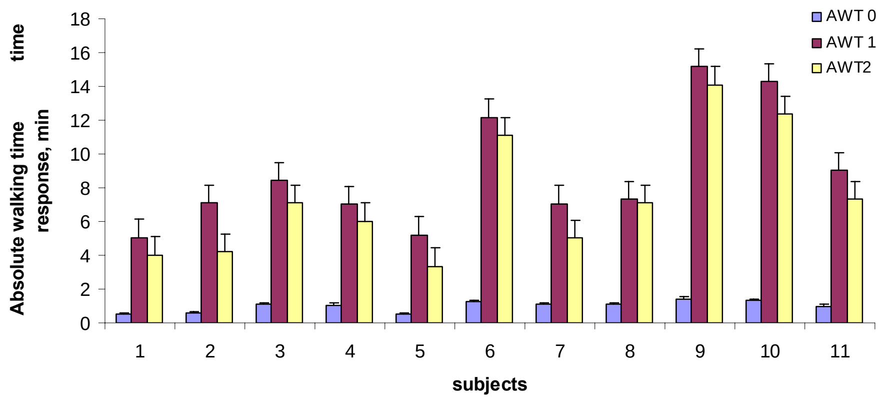


Figure 5.22 summarizes the Absolute Walking Time and Errors in a qualitative object detection and localization (number of heats). The time and mean error is represented for each ten subjects form the group A under two conditions: first trial and repeated trial. As mentioned before the data between subjects did not vary across conditions. Slight improvement in walking time were noted and for majority of subject a slight improvement on quality navigation.

(a)

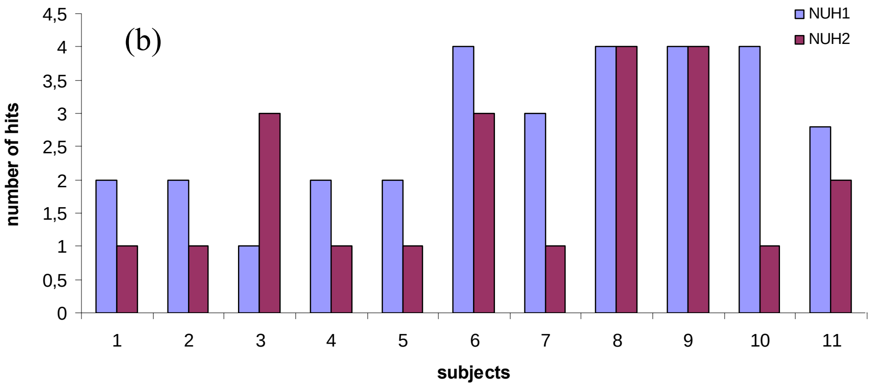


Figure 5.22 (a) Absolute walking time from Mobility Test I. The ten clusters of bars show results for the ten subjects from the group A as well as the mean across subjects. Results are shown for absolute walking time with the white cane (the blue colour), the absolute walking time for the first run (with magenta colour) and finally the absolute walking time for the second run (with yellow colour). The cluster of bars number 11 shows the mean time for ten subjects. (b) Errors from Mobility Test I group A are shown for the AWT1 and AWT2 for all subjects.

5.3.3.5 Conclusions

The experiments performed provide statistical results and first feedback regarding the validity of the accuracy of object detection in open environments. Despite the complexity of the task, the obtained results are great. The maximum mean Absolute Walking Time for the first and second trajectories is registered for the first group DBSV of 9,02min and 7,31min respectively. Also in the first group the walking accuracy is lower, the maximum number of hits and deviations from the trajectory is 7. Regarding the improvement on time and accuracy, we can mention that the slight improvements were detected for the repetition of the trajectory around 3min.

5.3.4 *Mobility Test II*

5.3.4.1 *Methods*

Mobility Test II is essentially an extension of the tests performed on Mobility Test I. The experiments consisted of a route leading through a popular and busy pedestrian area with a complex intersection. The test is divided into two parts. The part b) corresponds to the members of the group A, who carried out the experiment on a 145m long sidewalk close to the School for Blind. The part b) corresponds to the members of group B, who developed the experiment on a 500m path nearby the Institute. The data of the two groups were recorded and analyzed separately.

a) The first part of the 145 m long route was a traffic-calmed area without cars. Participants must travel the way trying not to hit the obstacles along the way, such as stone pillars, different types of street furniture, chairs, tables, other pedestrians, cyclists…, a fact which represented a constant challenge. The route then led to a complex intersection with four roads and a mid island which had to be crossed. There was no tactile pavement.

The scenario was set in a busy shopping area. There was a wide range of obstacles such as different kinds of street furniture, stalls, racks, differently sized advertising signs, cars parked on the pavement, bikes, huge stone pillars demarcating the first section of the area, steel bollards to keep cars away from the pedestrian area, but posing a real threat to a blind person. Test persons did not only have to cope with the many pedestrians bustling about. There were bikers buzzing around them. The video clips showed that not all sighted pedestrians were considerate towards the special needs of blind people.

b) Group B subjects were allowed to use the white cane with the condition as a further reference. The objective of the white cane was to detect the low obstacles and possible unevenness of the pavement. The scenario included columns of arcades, walls, thin and thick poles, a phone cabin threes and a newspaper kiosk. The experiment took the task only once.

5.3.4.2 *Mobility Test II Results*

Since the two groups carried out the experiment in different conditions the results are presented individually.

a) Surprisingly, taking all these aggravating circumstances into account, the blind test persons were obviously better able to cope with the challenges arising from this exercise in comparison with the Mobility Test I scenario. This is indicated, amongst others, by the fact that the difference between absolute walking time scores decreased. To give an example, the lowest score of AWT is 6,35 min to 5,07min for the Mobility Test I scenario. Thus the amount of time TP01 needed to travel the route of the scenario without the cane is six times higher for Mobility Test I and even nine times higher compared with the Laboratory scores while it dropped to just two times in Mobility Test II. Further evidence of this is supplied by taking a look at the average scores in the Table 5.6

Table 5.6 Mobility Test II results for the group A

Mobility Test II 1st Run										
	TP 01	**TP 02**	**TP 03**	**TP 04**	**TP 05**	**TP 06**	**TP 07**	**TP 08**	**TP 09**	**TP 10**	**mean**
AWT-1	3,26	2,21	4,57	4,08	3,1	5,5	2,29	4,55	7,07	6,16	4,31
AWT-2	7,01	9,21	12,09	11,14	6,35	16,45	8,02	10,18	20,19	19,41	12,09
NUH2	7	6	5	2	3	7	5	4	8	6	5,3
NOC2	2	5	4	3	2	4	3	4	4	3	3,4

Mobility Test II 2nd Run											
	TP 01	**TP 02**	**TP 03**	**TP 04**	**TP 05**	**TP 06**	**TP 07**	**TP 08**	**TP 09**	**TP 10**	**mean**
AWT-3	5,29	8,08	10,29	7,29	6,22	13,04	6,34	8,01	15,22	15,01	9,36
NUH3	4	3	5	4	2	5	4	3	6	5	4,1
NOC3	2	4	5	1	2	3	1	3	2	4	2,7

The findings reveal that there are some differences between individual blind test persons as indicated by the AWT range. To give an example, the respective AWT-2 score is 6,35 min to 20,19min, meaning that the amount of walking time spent by the slowest test person was three times that of the fastest. Again we might assume here, that there is a link to age, as the highest scores relate to the upper age band, though the range is not as prominent as regarding the Mobility Test II and the Laboratory exercises.

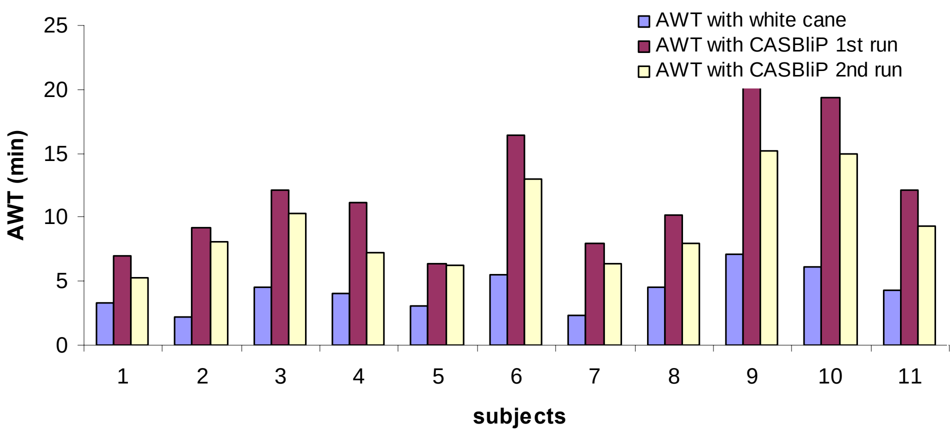


Figure 5.23 Absolute Walking Time results from Mobility Test II. The eleven clusters of bars show results for the ten subjects as well as the average across subjects.

More important is, however, that the role of training becomes evident again. After completing the first test run, the mobility instructor discussed with the test persons major problems and made proposals how the challenges of the route might be addressed more adequately. This obviously helped to improve the mobility performance as measured by a slight decrease in the relative walking time scores of the second test run.

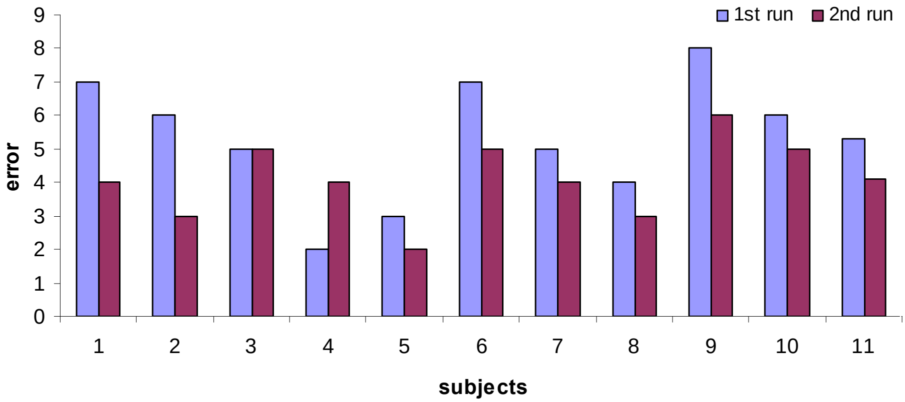


Figure 5.24 Number of hits results. The eleven clusters of bars represent the results for ten subjects and the average across subjects.

The findings reveal that test persons were able to avoid collisions with objects or people to a different extent. Some test persons had as many as 7 and 8 (near) collisions while others did considerably better. These differences are shown in Table 5.6**.** However, the role of training becomes clear again by comparing the NH scores from both test runs. In 8 out of 10 cases the number of unintended collisions had been reduced over against 2 remaining cases where no improvement was made or where even a change for the worse had been observed respectively. In addition improvement of mobility performance is confirmed by the NH average which dropped from 5 hits to 4 for the whole group after training.

The findings clearly show that all blind test persons were able to walk for longer distances without collisions after training.

Further evidence of the fact that training made a difference is supplied by taking a close look at the number of corrections needed to enable those blind test persons to continue their travel who had lost their bearing. The range of corrections was 2 to 5 in the first and 1 to 5 in the second test run. Thus no blind test person was able to do the exercise without corrections. By comparing the results from both test runs it becomes evident that in 6 out of 10 cases improvement hat been made, while in 2 cases no change was observed and in 2 further cases the number of correction had even risen. This is confirmed by looking at the total number of corrections which decreased from 34 to 27 after training.

b) The results of the group B are shown in Table 5.7. From these test we can see that beside that there were less obstacles on the route, the subject TP2 had 12 hits, which means that he had difficulties in orientation. The route was located beside a wall, a fact that means that almost all subjects used the wall as reference. Almost all subjects took advantage of the wall, avoiding the moving objects.

Table 5.7 Mobility Test II results for the group B

Mobility Test II 1st Run											
	TP 01	TP 02	TP 03	TP 04	TP 05	TP 06	TP 07	TP 08	TP 09	TP 10	mean
AWT-4	17,00	12,00	16,00	17,54	13,20	9,30	10,20	14,40	14,10	17,50	14,16
NUH4	2	12	1	1	4	0	0	0	0	0	2

5.3.4.3 Mobility Test II Conclusions

These results showed that the number of hits is related to the presence of the object and subject velocity. Subjects accurately estimated the object presence as the acoustical interpretation. Performances in the decontrolled condition confirmed again the well known notion of practice, it means that an accurate travel can be achieved when the subject is relaxed and tested the system during a period. In the case of mobility through unknown environment is always quite difficult both for blind and people with normal vision. Unknown environment always requires more attention, concentration and longer time in analyzing the surrounding environment. Beside that the environmental noises interfere on the human hearing. In these situations the subjects must be double concentrated, to separate the street lighting sounds, the human noises etc.

5.4 General conclusions of the navigation tests

The findings of the current study were consistent with those of previous studies. There are several possibilities to explain the decline in performances. One option is that not all subjects experimented electronic travel aids for blind people. There is small number of subject which has access to test different ETA systems and mainly acoustical systems, due to the limited projects in this area and the limitation of centres for blind involved on these projects.

Another option is the limitation on the present system usage. There were a small number of devices developed, and the subjects were not allowed to take them at home and test them during a long time.

One of the striking features of the data collected in these experiments is the presence of strong individual differences. The subjects varied in their baseline localisation accuracy and their vulnerability to acoustical sounds. A likely explanation for the differences noticed with acoustical sounds is that some subjects are more reliant with high tone sounds for localization than others. This means that the same volume for different subjects will represent different distances. People with hearing problems are also included in that situation. Also we can add the difficulties when localizing the objects in presence of surrounding noises.

A very interesting observation can be reached by comparing the results from the three experiments described above. In experiment Basic Learning Protocol, it was found that subjects had great difficulties on sound externalization. These are presumably related to the fact that the subjects must perceive the difference between the monaural and binaural cues of the sounds. These means that the perceived sounds through headphones must be interpreted as coming from the surrounding.

In experiment Laboratory Test there were many instances in which different objects sounded as one object due to the near distance between or angle of vision of the subject. Generally, the subjects achieved great results. Subjects were asked to localize and walk through a scenario in these experiments and subject performed well. Taken together, these results are surprising, the subjects were able to localize the objects in pairs, to choose the free path between the columns and pass, but they were not able to see the difference between very near objects. Sometimes, the subjects lost the trajectory direction walking to a wrong direction. This discrepancy is interesting and it deserves some considerations. For example, during the surrounding exploration, when in the area of vision no sounds were perceived, this meant that no objects were present in that area; in that case the subjects must return back and look for objects. A possible explanation for the loss of ability in object localization is that subjects did not have enough time to test the system.

Having a look at the experiment Mobility Test I and Mobility Test II, it becomes evident that whilst the group in Mobility Test I was able to walk without colliding with an object for roughly 15 meters, after training the group performed much better in the Mobility Test II trials, in which the respective distance had increased to 35 meters and more. As Mobility Test II followed Mobility Test I, it can be assumed that test persons went into the trials more experienced and better trained. This might be considered as a further indication of the important role of appropriate training.

An alternative interpretation is that location processing was not the limiting factor, but that subjects were unable to effectively perceive all stimuli in the navigation task. There are two possible realisation of this hypothesis. Firstly subjects may adopt a divided strategy, and attempt to monitor the different locations simultaneously. On the other hand, it may be that subjects adopt a selective attention strategy, where the attention is paid in a single object and then the attention is switched to the other. In this case, it may be that duration is the limiting factor. It may take a prolonged time to effectively concentrate and switch attention along the navigation task.

6. CHAPTER VI: CONCLUSSIONS

CHAPTER VI: CONCLUSSIONS

1. The representation of 3D objects using sounds positioned in a single line with HRTF functions can be used both in indoor and outdoor environments for representing the world information to totally blind people. As it has been shown in the first chapter, none of the Electronic Travel Aids developed up to date fulfils all criteria for successful mobility of blind people. With this idea, the present thesis demonstrated the possibility of development of a cognitive electronic travel aid able to transform environmental information into acoustic signals helping blind people to detect the obstacles in the front of view and to take decisions on selecting their walking trajectory.
2. It has been also proven that with the developed system, the blind people is able to travel confidently and safety. Due to the fact that the system gives information on the environment comprised between 0 and 15m within a range of 32º relative to the right and left side of the user, the system gives more information than the white cane or the existing ETA systems, which are constricted to distances up to 6m.
3. A system based on a CMOS Time of Flight laser is used for detection of the objects between 1 and 5m, giving a measurement error lower than 1% distance for 100% target. Also the system can be used with decreasing resolution and accuracy in ~~a~~ distances over 5 meters.
4. An object detection system based on Stereo Cameras can be used in parallel or individually. It shows that it is possible to represent the real environment using the depth map method extracted from the stereo vision, to extract the objects, to classify them and to extract the free paths.
5. It has been also proven that using spatial acoustic sounds, the processing and delivering time is not long, since the used spatial sound has 2048 samples. Due to the short acoustic sounds used on the system, the representation of the objects detected by the system is delivered in real time. It is not necessary to inform the user about the objects located in front of view one by one; the system represents the objects at different peach, depending on their distance and type, thus the user perceives the whole image of the environment being able to take decisions.
6. It has been shown that non-individual HRTFs produce the externalization effect. The spatial acoustic sounds delivered by the

headphones, are perceived by the user brain as sounds produced by the environment.

7. Also, it is analysed how the time difference or the inter-click interval plays an important role on sound source localization.
8. In the thesis, the maximum and minimum inter-click interval threshold for an accurate localization are also analyzed. In Chapter 4, it is proven that at the ICI of 200ms a saltation phenomenon is perceived, and at ICI's shorter than 10ms, the sounds are perceived as a blur.
9. Chapter 4 also shows that it is possible to localize a moving sound source with the spatial sounds.
10. It is proven than the spatial sounds used by the system do not interfere with the environmental cues, such as speech, car and/or street noises.
11. To sum up, the thesis shows the system usability both in indoor and outdoor situations, because the CMOS Time of Flight laser is especially suitable for all these conditions (closed environment, darkness, with light, for white objects or walls, etc...).

7. CHAPTER VII: FUTURE WORK

CHAPTER VII: FUTURE WORK

The system outcome of this thesis is a novel Electronic Travel Aid for blind people that collect the environmental information using the CMOS Time of Flight laser, process them and present to the user via acoustic sounds. Out coming from the experimental results presented in the previous chapters which provide some early statistical results and first feedback regarding the validity of the prototype/system design and functionality, several future improvements may be investigated:

1. to miniaturize the whole electronics of the system
2. to reduce the electronics and equipments
3. to improve the image processing algorithms
4. to improve the object identification accuracy
5. to improve free paths identification accuracy
6. to improve the travel speed and time
7. to implement into the system a vision system for partially blind people
8. to define optimum distance between two consecutive objects sounds
9. to improve the acoustic sounds
10. to improve the independent mobility with the CASBliP system

As has been emphasized through this thesis, complex environments containing concurrent sound sources present interesting challenges to the auditory system, and particularly to the process underlying spatial hearing. The experiments with the navigation system and psychoacoustic ones described in this thesis provide us with many clues as to how the nervous system might represent sound sources and how it create a virtual visual image of the detected environment through acoustical signals. However, complementary psychological investigations are vital if we are test the validity of these speculations. The appearance of several recent publications provides encouragement in this area.

One of the most important roles of the auditory system is to monitor the environment and specially to detect changes in the surrounding environment. In terms of perceiving the spatial layout of the environment, auditory information is an extremely important component to visual information.

Sound localization has contributed enormous into our current understanding of mechanisms for object detection and localization in navigation system.

In this sense, is essential to continue the investigation on sound localization, using different methods and new technologies. Among them several tasks must to be investigated:

11. To develop experiments on sound localization in anechoic chamber. The signals to be measured directly on the human auditory system.
12. To make an analysis of how the interaural differences (interaural time difference and interaural level difference) are pass to localization decisions on the human auditory system, under the condition than the interaural parameters get distortioned due to the camera reflections.
13. To study new methods of the sound source localization improvement when the subject is moving.
14. To measure the sound field which is perceived by the subjects when are moving.
15. to measure the sound localization directly on the human auditory system for different level of disorders
16. to make an analysis of the interaural differences as azimuth function

It is hoped that further efforts to characterize and analyze sound localization will clarify some of the issues and doubts and finally, will increase our understanding of the role of auditory spatial representations in natural environments.

8. CHAPTER VIII: REFERENCES

CHAPTER VIII: REFERENCES

Al`tman Ya.A., Gurfinkel V.S., Varyagina O.V., Levik Yu.S. (2005). “The effect of moving sound images on postural responses and the head rotation illusion in humans,” Neuroscience and Behavioral Physiology, 35 (1), 103-106

Amedi A., Camprodon J., Bermpohl F., Fox S., Merabet L., Meijer P., Pascual-Leone A.. (2004). Visual-to-auditory sensory substitution in patient blind user: Neural correlates and potential application in neuroprostheses research. poster

Armstrong M., Zisserman A. and Hartley R. (1996). “Euclidean Reconstruction from Image Triplets", *Computer Vision - ECCV'96*, Lecture Notes in Computer Science, Vol. 1064, Springer-Verlag, pp. 3-16

Ando B., (2003). Electronic Sensory systems for the Visually Impaired. IEEE Instrumentation & Management Magazine, pp. 62-67

Angell J.R., Fire W., The monaural localization of sound, Psychol Rev. 1901; 8, pp.225-243

Action for Blind People. Easy Walk – a new system for people with sight loss. http://www.actionforblindpeople.org.uk

Bach-y-Rita P., Collins C. C., Saunders F. A., White B., Scadden L., (1969). Vision Substitution by Tactile Image Projection. Nature, Vol. 221, pp.963-964

Benham T. A. (1954). Guidance Device for the Blind. Phys. Today, 7, No. 12, pp. 11-14

Benjamin J. M. (1968). A review of the veterans administration blind guidance project. Bull. Of Prosth. Res., pp. 63-90

Bentzen B., P. Mitchell (1993). Audible signage as a wayfinding aid: comparison of verbal landmark and talking signs. Draft Report to the American Council of the Blind

Blauert J. (1997). “Spatial Hearing: The Psychophysics of Human Sound Localization,” revised edn, The MI Press, Cambridge, MA, USA,

Book (1986) “Electronic travel aids: new directions for research,” National Research Council, Chap.6, 67-90

Bondyfalat D. and Bougnoux S., (1998). “Imposing Euclidean Constraints During Self-Calibration Processes", *Proc. SMILE Workshop (post-*

ECCV'98), Lecture Notes in Computer Science, Vol. 1506, Springer-Verlag, pp.224-235

Bourbakis N., Kavraki D.. (1996, Nov.4-5). Intelligent Assistants for handicapped people's independence: case study. IEEE International Joint Symposia on Intelligence and Systems, Rockville, MD, p. 337-344.

Bourbakis N. (2008). Sensing surrounding 3-D space for navigation of the blind - A prototype system featuring vibration arrays and data fusion provides a near real-time feedback. IEEE Engineering in Medicine and Biology Magazine, v. 27, Issue 1, pp. 49-55.

Bourbakis N., Kakumanu P.. (2008). Skin-based Face Detection- Extraction and Recognition of Facial Expressions. In Applied Pattern Recognition, Vol.91, (book chapter).

Brabyn J. (1982). New Developments in Mobility and Orientation Aids for the Blind. IEEE Transactions on Medical Engineering, vol BME-29, NO 4, pp. 285-289

Brabyn J.A. “A review of mobility aids and means of assessment”, in D.H.Warren & E.R.Strelow edition “Electronic spatial sensing for the blind”, Boston: Martinus Nijhoff, pp.13-27

Brabyn J. (1997). Technology as a support system for orientation and mobility. American Rehabilitation

Bruce H., Deatherage and Ira J. Hirsh, Auditory Localization of Clicks, Journal of the Acoustical Society of America 1959; 31(4), pp. 486-492

Bruce V., Green P. R., Georgeson M. A. “Visual Percption: Physiology, Psychology and Ecology”, Psychology Pres, p5-14, 1996Butler R.A., Humanski R.A., Musicant A.D., Binaural and monaural localization of sound in two-dimensional space, Perception 1990, 19, pp. 241/256

Capp M., Picton P. (2000). The optophone: an electronic blind aid. Engineering Science and Education Journal, pp. 137-143

CIDAT. Manual de usuario del Systema Tormes. CIDAT, ONCE

Collins C. C. (1985). On mobility aids for the blind. Eds. D. H. Warren & E. R. Strelow, Electronic spatial sensing for the blind, pp. 35-64

Costa G., Gusberti A., Graffigna J. P., Guzzo M., Nasisi O. (2007). Mobility and Orientation aid for blind persons using Artificial Vision. Journal of Physics

Dakopoulos D., Bourbakis N. (2008, July 15-19). Preserving Visual Information in Low Resolution Images During Navigation of Blind. 1st

International Conference on PErvasive Technologies Related to Assistive Environments, Athens, Greece.

Dakopoulos D., Bourbakis N. (2007, Oct. 15-17). A 2D Vibration Array as an Assistive Device for Visually Impaired. IEEE Int. Conf. on BIBE07, Boston, MA, vol. II, pp. 930-937.

Debnath N., Tangiah J. B., Pararasaingam S., A. S. A. Kader, (2004). A mobility aid for the blind with discrete distance indicator and hanging object detection. TENCON IEEE, pp. 664-667

Denham J., Leventhal J., McComas H. "Getting from point A to point B: A review of two GPS systems", AccessWorld, 5,6 2006

Dodson A.H., More T., Moon G.V "A navigation system for the blind pedestrian", Proceedings of GNSS 99, 3rd European Symposium on Global Navigation Satellite Systems, pp. 513-518, 1999

Douglas S. B, W. R. Raibiowitz., Auditory Localization in the Near-Field, Journal of the Acoustical Society of America 1999; 106 (3): pp.1465-1479

Douglas S. B, Nathaniel I., W. R. Raibiowitz. Auditory localization of nearby sources II. Localization of a broadband source, Journal of the Acoustical Society of America 1999; 106 (4), pp.1956-1968

Douglas G., Tobin M., Bozic N., Hill E., "MOBIC: A SYSTEM FOR FACILITATING INDEPENDENT MOBILITY AND NAVIGATION FOR BLIND PEOPLE" Proceedings of ICEVI's Xth World Conference, Sao Paolo, Brazil, 3-8 August 1997

Dufour, O. Despres, Candas V. "Enhanced sensitivity to echo cues in blind subjects", Exp. Brain Res, Springer, 165, pp. 515-519, 2005

Dunai L., Peris F G.., Defez B. G., Ortigosa A.N., Brusola S F. (2009). "Perception of the sound source position", Applied Physics Journal, (3), 448-451

Dunai L., Peris F G.., Defez B. G., Ortigosa A.N., (2009). "Acoustical Navigation Sysyem for Visual Impaired People", LivingAll European Conference

Dupress J. K. (1963). The Requirements for Successful Travel by the Blind. In L. L. Clark Eds. 1, pp. 7-11

Espinosa M.A., Ungar S., Ochaita S., Blades M., Spencer C. "Comparing methods for introducing blind and visual impaired people to unfamiliar urban environment", Journal of Environmental Psychology, 18 pp. 277-287, 1998

ESA. (2008, October). Space technology to help the blind. ESA navigation-news

ESA. (2003, June). More autonomy for blind people thanks to satellite navigation. ESA news

Fancy R., Leroux R., Jucha A., (2006). Electronic travel aids and electronic orientation aids for blind people: Technical, rehabilitation and everyday life points of view. CVHI 2006 conference

Farmer L.W. "Mobility devices", In L. W. Welsh & B.B. Blasch (Eds.), Foundations of orientation and mobility, pp. 357-412, 1980. New York> American Foundation for the Blind

Farmer L., Smith D. (1997). Adaptive Technology. Book chapter, Foundations on Orientation and Mobility, Eds. 2, pp.231-260

Farmer W. (1978). Mobility devices[a]. Bulletin of Prosthetic Research, pp-47-118

Farmer L. W., (1975). Travel in adverse weather using electronic mobility guidance devices. The new outlook for the blind, 69 (10), pp. 433-439

Farcy R., Leroux R., Jucha A., Damaschini R., Gregoire C., Zogaghi A. "Electronic travel aids and electronic orientation aids for blind people: Technical, rehabilitation and everyday life points of view", CVHI Conference, 2006

Fay R.R., Popper A.N. (2005) Introduction to sound source localization, Springer Handbook of Auditory research, sound source localization, pp. 1-5

Fay R.R. (2005). Sound source localization by fishes. Springer Handbook of Auditory research, sound source localization, pp. 36-66

Fornoz A. P., Sommerbalde R.J., Rappaz B., Safran B.A., M. Pelizzone. (2005). Simulation of Artificial Vision, III: Do the Spatial or Temporal Characteristics of Stimulus Pixelization Really Matter?. Investigative Ophthalmology & Visual Science, Vol. 46, No. 10, pp. 3906-3912

Foulk E. "The perceptual basis for mobility", AFB Res. Bull., vol 23, pp. 1-8, 1971

Freiberger H. A. M. (1974). Mobility Aids for the Blind. Bulletin of Prosthetic Reseatch, pp. 73-78

Golledge R.G., Klatzky R.L., Loomis J.M., Speigle J., Tietz J. "A geographical information system for GPS based personal guidance system", International Journal of Geographical Information Science, Vol.12, No.7, pp.727-749, 1998

Goledge R. G. "Tactual strip maps as navigational aids", Journal of Visual Impairment & Blindness, 85, pp.296-301, 1991

Hari R., Illusory directional hearing in humans, Neuroscience Letters 1995; 189, pp.29-30

Harris G.G. Binaural interactions of impulsive stimuli and pure tones. Journal of the Acoustical Society of America 1960; 32, pp.685-692

Harris J. G., Pu Ch.J., Principe J.C., (2000). "A monaural cue sound localizer," Analog Integrated Circuits and Signal Processing, 23, 163-172

Hartmann W.M., Wittenberg A., (1996). "On the externalization of sound images," J. Acoust. Soc.Am. 99 (6): 3678-3688

Hartley R., (1994). "Euclidean reconstruction from uncalibrated views", in : J.L. Mundy, A. Zisserman, and D. Forsyth (eds.), *Applications of Invariance in Computer Vision*, Lecture Notes in Computer Science, Vol. 825, Springer-Verlag, pp. 237-256

Hoyle B. S. (2003). The Bathcane – mobility aid for the vision impaired and the blind. IEE Symposium on Assistive Technology, pp. 18-22

Hub A. (2008). Guiding grids in augmented maps for precise installation- free worldwide blind navigation. Conference Proceedings of CSUN Conference.

Hub, J. Diepstraten and T. Ertl. Design and Development of an Indoor Navigation and Object Identification System for the Blind. Proceedings of the ACM SIGACCESS conference on Computers and accessibility, Designing for accessibility, Atlanta, GA, USA, 147-152, 2004.

Hub, S. Kombrink, K. Bosse and T. Ertl. TANIA - A Tactile-Acoustical Navigation and Information Assistant for the 2007 CSUN Conference. To appear in: Conference Proceedings of the California State University, Northridge Center on Disabilities' 22nd Annual International Technology and Persons with Disabilities Conference, March 19-24, Los Angeles, CA, USA, 2007.

HumanWare "Braile Note GPS", www.humanware.com

Jeffress L. A. A place theory of sound localization. *J. Comp. Physiol. Psychol.* 61: 468-486, 1948.

Joana C. Kidd, John H. Hogben, Quantifying the auditory saltation illusion: An objective psychophysical methodology. Journal of the Acoustical Society of America 2004; 116 (2) pp.1116-1125

Junius D., Riedel H., Kollmeier B., (2007). “The influence of externalization and spatial cues on the generation of auditory brainstem responses and middle latency responses,” Hearing Research 225, 91-104

Kawai Y., Fumiaki T. “A support system for visually impaired persons to understand three-dimensional visual information using acoustic interface”, IEEE transaction, pp. 974-977, 2002

Kay L. (1964). An ultrasonic sensing probe as an aid to the blind. Ultrasonics, Vol. 2, No. 53

Kelly G.W. (1981). Sonic Orientation and Navigational Aid (SONA). Bulletin of Prosthetics Research, 1, 189.

Kim H.Y., Suzuki Y, Sh. Takane, Sone T. (2001). “Control of auditory distance based on the auditory parallax model,” Applied Acoustics 62, 245-270

Kistler D.J., Wightman F.L., (1996). “A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction a, b), 1992,” J. Acoust. Soc.Am. 91 (3), 1637-1647

Klatzky R. L., Marston J. R., Guidice N. A., Golledge R. G., Loomis J. M. “Cognitive load of navigating without vision when guided by virtual sound versus spatial language”, Journal of Experimental Psychology: Applied, Vol. 12, No. 4, pp. 223-232, 2006

Kramer G., (1994). Auditory Display: Sonification, Audition and Auditory Interfaces. Proceedings Volume XVIII, Reading MA: Addition Wesley

Kulkani A., Colburn S.H., (1998). “Role of spectral detail in sound-source localization,” Nature, 396, 747-749

Kuc R. (2002). Binaural Sonar Electronic Travel Aid Provides Vibrotactile Cues for Landmark, Reflector Motion and Surface Texture Classification. IEEE Transactions on Biomedical Engineering, Vol. 49, No 10, pp. 1173-1180

Lagace Y., (April 2005). HumanWare`s Trekker GPS system creates handheld personal guide with HP ¡PAQ Pocket PC. HP notice

Liebowitz D. and Zisserman A. (1997). ”Combining Scene and Auto-calibration Constraints", *Proc. ICCV*, pp.293-300, 1999Litovsky, Ruth Y., Brad Rakerd, Tom C. T. Yin, and William M. Hartmann. Psychophysical and physiological evidence for a precedence effect in the median sagittal plane. *J. Neurophysiol.* 77: 2223-2226

Loomis J.M., Golledge R.G., Klatzky R.L. "Navigation system for the Blind: Auditory display modes and guidance", Journal of Presence, Vol.7, No.2 pp.193-203, 1998

Loomis J.M., Klatzky R.L., Reginald G., Golledge R.G. "Navigating without Vision: Basic and Applied Research", Optometry and Vision Science, Vol.78, No.5, pp. 282-289, 2001

Loomis J.M., Hebert C., Cicinelli J.G., (1990). "Active localization of virtual sounds", J. Acoust. Sot. Am., 88, 4, pp.1757-1764

Loomis, J. M., Golledge, R. G., Klatzky, R. L., & Marston, J. R. (2006). Assisting wayfinding in visually impaired travelers. In G. Allen (Ed.), Applied spatial cognition: From research to cognitive technology. Mahwah, N. J.: Lawrence Erlbaum Associates.

Loomis, J. M., Marston, J. R., Golledge, R. G., & Klatzky, R. L. (2005). Personal guidance system for people with visual impairment: A comparison of spatial displays for route guidance. Journal of Visual Impairment & Blindness, 99*,* 219-232

Loomis J. M., Golledge R. G., Klatzky R. L. (2001). GPS-Based Navigation Systems for the Visually Impaired. Pp. 429-446

Makous J.C., Middlebrooks J C. "Two-dimensional sound localization by human listeners", J. Acoustt. Soc. Am. 87 (5), pp.2188-2200, 1990

Makous J. C., Middlebrooks J C. (1990). Two-dimensional sound localization by human listeners". J. Acoustt. Soc. Am. 87 (5), pp.2188-2200

Mann R.W., (1970). Mobility aids for the blind—An argument for a computer-based, man-device-environment, interactive, simulation system. Pro Conf on Evaluation of Mobility Aids for the Blind, Washington, DC: Com. on Interplay of Eng. with Biology and Medicine, National Academy of Engineering. pp. 101–16.

Meijer P. B. L. (2005). A Modular Sysnthetic Vision and Navigation System for the Totally Blind. World Congress Proposals

Meijer P. B. L. (1992). An experimental system for auditory image representations. IEEE Transaction Biomedical Engineering, BME-39, (2), pp. 112-121

Moons T., Van Gool L., M. Proesmans and Pauwels E., (1996). "Affine reconstruction from perspective image pairs with a relative object-camera translation in between", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no.1, pp. 77-83

Mora J.L. G., Rodriguez-Hernandez A.F., Martin F., Castellano M.A. "Seeing the world by hearing: Virtual Acoustic Space (VAS) a new space perception system for blind people", Proceedings of the 2nd Information and Communication Technologies Conference, ICTTA'06, IEEE pp. 837-842, 2006

Morales T. M., Berrocal A. M.. (2005, Octomber). El GPS como sistema de ayuda a la orientacion de personas ciegas. IIICV-INTEREVISUAL Autonomia Personal

Morrissette D. L. Goddrich G.L., Henesey J.J., (1971) "A follow-up-study of the Mowat sensors applications, frequency of use and maintenance reliability." J. Vis. Impairment Blidness, Vol. 75, 6, 244-247

NRC (1972), The cane as a mobility aid for the blind. A report of a conference. September 10-11, 1971

Nokia forum. Open Source Software Helps to Turn a Smartphone into Talking GPS.

Ortigosa Araque N., Dunai L., Peris Fajarnes G., Santiago Praderas V., Dunai I. (2009, 11 December). Sound map generation for a prototype blind mobility system using multiple sensors. ABLETECH'08 Conference

Passini R., Prolux G. "Wayfinding without vision: An experiment with congenitally totally blind people", Environment and Behaviour, 20, pp.227-252, 1988

Petrie H., Johnson V., Strothotte T., Raab A., Michel R., Reichert L., Schalt A. (1997). MoBIC: An Aid to Increase the Independent Mobility of Blind Travellers. British Journal of Visual Impairment, Vol. 15, No. 2, pp. 63-66

Pollefeys M., Van Gool L. and Oosterlinck A. (1996). "The Modulus Constraint: A New Constraint for Self-Calibration", *Proc. 13th International Conference on Pattern Recognition*, IEEE Computer Soc. Press, pp. 349-353

Russell L. (1965). Travel Path Sounder. Proceedings, Rotterdam Mobility Research Conference. N.Y: American Foundation for the Blind

Sainarayanan G., Nagarajan R., and Yaacob S., (Jan. 2007). Fuzzy image processing scheme for autonomous navigation of human blind‛, Appl. Soft Comput. 7, 1 pp. 257-264.

Santiago Praderas V., Ortigosa N., Dunai L., Peris Fajarnes G. (2009, 10-12 June). Cognitive Aid System for Blind People (CASBliP). Proceedings of XXI Ingehraf-XVII ADM Congress p. 31

Sato E., Yamaguchi T., Harashima F. (2007, April). Natural Interface Using Pointing Behavior for Human-Robot Gestural Interaction. *IEEE Trans. on Industrial Electronics,* vol. 54, no. 2, pp. 1105-1112

Shoval S., Ulrich I., Borrenstein J. (2000). Computerized obstacle avoidance systems for the blind and visually impaired" Invited chapter in "Intelligent systems and technologies in rehabilitation engineering", Editors Teodorescu, H.N.L & Jain, L.C.,CRC, pp. 414-448

Starkiewicz W., Kuliszewski T., (1963). Active energy radiating system: the 80-channel Electroftalm. Proceedings Int. Congress on Technology and Blindness, Vol. 1, pp. 157-166

Strothotte T., Petrie H., Johnson V., Reichert L. (1997). MoBIC: user needs and preliminary design for a mobility aid for blind and elderly travellers. British Journal of Visual Impairment, Vol. 15, No. 2, pp. 63-66

Strutt J.W. (1907). "On our perception of sound direction," Philos. Mag.; 13, 214.232

Takahashi T. T., Keller C. H. "Representation of multiple sound sources in the owl's auditory space map", The Journal of Neuroscience, 14(8), pp. 4780-4793, 1994

The Loadstone GPS team. (2008, November). Open Source Software Helps to Turn a Smartphone into Talking GPS. http://www.loadstone-gps.com

Vadakkepat P., De Silva L.C., Jing L, Ling L.L.. (2008, March). Multimodal Approach to Human-Face Detection and Tracking. *IEEE Trans. on Industrial Electronics,* vol. 55, no. 3, pp. 1385-1393

Versenyi G. (2007). "Localization in a Head-Related Transfer Function-based virtual audio sysnthesis using additional high-pass and low-pass filtering of sound sources," Acoust. Science & Technology, 28 (4), 244-250

Walker B. N., Lindsay J. (2006). Navigation performance with a virtual auditory display: Effects of beacon sound, capture radius, and practice. Human Factors, Vol. 48, No. 2, pp. 265-278

Walker B. N., Lindsay J. (2005). Using virtual environments to prototype auditory navigation display. Asst. Technology, 17, pp. 72-81

Walker B.N., Lindsay J. (2004). Auditory Navigation Performance is Affected by Waypoint Capture Radius", Proceedings of ICAD 04, Sydney, Australia

Wenzel E., Arruda M., Kistler D., Foster S. (1993). "Localization using non-individualized head-related transfer functions," J. Acoust. Soc.Am. 94, 111-123

Wenzel E.M., (1992). "Localization in virtual acoustic display," Presence Telop Virt. Environ. 1, 80-107

Wightman F.L., Kistler D.J., Monaural sound localization revisited, J. Acoust. Soc. Am., 1997; 101, pp.1050-1063

Wilson J., Walker B. N., Lindsay J., Cambias C., Dellaert F. (2007), 11th IEEE International Symposium on, pp. 91-98

WBU (2009). White cane safety day. World Blind Union, Press release, October, Canada

WBU (2009). White cane safety day. World Blind Union, Press release, October, Canada

Leung J., Carlile S., (2004). Spatialisation of Talkers and the Segregation of Concurrent Speech. Association for Research in Otolaryngology.