



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA DEL DISEÑO

Tesis Doctoral

Desarrollo de un sistema de entrenamiento
acústicovirtual para localización de sonidos
virtuales para personas invidentes

Valencia, marzo de 2014

Autor: D. Ismael Lengua Lengua

Director: Dr. D. Guillermo Peris Fajarnés
Codirectora: Dra. Dña. Larisa Dunai Dunai

ÍNDICE GENERAL

DEDICATORIAS	15
AGRADECIMIENTOS	17
RESUMEN.....	19
RESUM	21
ABSTRACT.....	23

PARTE I INTRODUCCIÓN - OBJETIVOS

1. INTRODUCCIÓN	25
1.1 Generalidades.....	27
1.2 HRTF	37
1.3 3D-CMOS – Time of Flight.....	41
1.4 Publicaciones científicas y registro propiedad intelectual.....	45
1.5 Referencias.....	48
2. OBJETIVOS.....	51

PARTE II PUBLICACIONES

3. ARTÍCULO I..... 53

Ismael Lengua, Larisa Dunai, Guillermo Peris-Fajarnés, Beatriz Defez.

DISPOSITIVO DE NAVEGACIÓN PARA PERSONAS INVIDENTES BASADO EN LA TECNOLOGÍA TIME OF FLIGHT.

Dyna, año 80, Edición 179, pp. 33-41, Medellín, junio, 2013. ISSN 0012-73. 53

Abstract..... 53

3.1 Introducción 54

3.2 Arquitectura del dispositivo 57

3.3 El sistema láser y el método de medida de las distancias..... 58

3.4 Sistema acústico y método de producción de distancias mediante sonidos..... 62

3.5 Resultados 66

3.6 Conclusiones 68

3.7 Agradecimientos..... 69

3.8 Referencias..... 69

4. ARTICULO II 73

Larisa Dunai, Guillermo Peris-Fajarnes, Ismael Lengua, Ignacio Tortajada Montañana.

MATERIAL SOUND SOURCE LOCALIZATION THROUGH HEADPHONES.

Acoustical Physics, 2012, Vol. 58, No. 5, pp. 610–617. ISSN 1063-7710..... 73

Abstract	73
4.1 Introduction.....	74
4.2 Listeners.....	78
4.3 Method.....	79
4.4 Results.....	81
4.5 Conclusions.....	87
4.6 Acknowledgments.....	87
4.7 References.....	87
5. ARTICULO III	89
Ismael Lengua, Larisa Dunai, Guillermo Peris-Fajarnés, Fernando Brusola.	
LOCALIZACIÓN DE SONIDOS EN ENTORNO ABIERTO POR PERSONAS CIEGAS.	
<i>Revista Acústica de la Sociedad Española de Acústica.</i> Vol. 45. Núm 3-4 2013.	
I.S.S.N. 0210-3680.	89
Abstract	89
5.1 Introduction.....	89
5.2 Descripción de los equipos.....	91
5.3 Descripción de la prueba.....	97
5.4 Resultados	98
5.5 Conclusiones	99
5.6 Referencias.....	100

6. ARTICULO IV101

Larisa Dunai, Beatriz Defez, Ismael Lengua, Guillermo Peris-Fajarnes.

3D CMOS SENSOR BASED ACOUSTIC OBJECT DETECTION AND NAVIGATION SYSTEM FOR BLIND PEOPLE.

IECON 2012 - 38th Annual Conference on IEEE Industrial Electronics Society. 4208-4215

Montreal, QC Date of Conference: 25-28 Oct. 2012,

ISBN: 978-1-4673-2419-9. 101

Abstract..... 101

6.1 Introduction 101

6.2 Description of the Acoustic System 104

6.3 Implementation of the sensor module 106

6.4 Implementation of the Acoustic Module 108

6.5 Experimental Results..... 111

6.6 Conclusions 117

6.7 Acknowledgment 119

6.8 References 119

PARTE III DISCUSIÓN GENERAL DE RESULTADOS

7. DISCUSIÓN GENERAL DE RESULTADOS.....125

7.1 Referencias 127

PARTE IV CONCLUSIONES

8. CONCLUSIONES129

ÍNDICE DE FIGURAS

Figura 1 Método de medición por Tiempo de Vuelo (ToF)	19
Figura 1.1 La población mundial con discapacidad visual. (Fuente: Organización Mundial de la Salud)	27
Figura 1.2 Estimación del crecimiento de la población mundial con problemas de visión (Fuente: Organización Mundial de la Salud, World Blind Union).....	28
Figura 1.3 Principales causas de ceguera (Fuente: Organización Mundial de la Salud).....	29
Figura 1.4 La cámara anecoica con bajo nivel de reverberación y sistema robótico. Las paredes y el suelo de la sala están cubiertas con cuñas que absorben el sonido. El sistema robótico conjunto puede ser movido para colocar el altavoz en cualquier ubicación en radio de 5 metros. Una silla para el maniquí se representa en una plataforma fija.....	38
Figura 1.5 Imagen real con los elementos del sistema: gafas (elemento sensor 1x64 pixeles del láser miniaturizado e integrado).	41
Figura 1.6 Coche de prueba con una cámara PMD para seguridad de los peatones y la detección de Pre-colisión, respectivamente [5].....	42
Figura 3.1 Imagen de Cognitive Aid System for Blind People	58
Figura 3.2 Principio de cálculo del área de sensibilidad del píxel y el diagrama de tiempo del método de medida de distancias Time-Of-Flight.	59
Figura 3.3 Circuito del píxel.....	60
Figura 3.4 Diagrama de tiempo de los siete ejercicios de la prueba de aprendizaje. El eje x representa cada uno de los siete ejercicios; los valores del eje y representan el tiempo en minutos; las barras en colores representan las tres pruebas realizadas para cada ejercicio.	66

Figura 3.5 El diagrama de tiempo de las 5 pruebas de laboratorio. El eje x representa cada una de cinco pruebas (la prueba 1 representa el resultado con ayuda del bastón, las siguientes dos columnas representan los resultados promedio de ida y vuelta del ejercicio de los diez participantes de Alemania, mientras que las columnas 4 y 5 representan el promedio de los resultados de ida y vuelta del grupo de Italia). Los valores del eje y representa el tiempo en minutos.67

Figure 4.1 A convolution representation of the two sequences. In (a), it is shown the mathematical representation of the convolution, in (b) the convolution algorithm between the HRTF and the previously generated signal is presented.76

Figure 4.2 Representation of sound waves obtained after the convolution between the bongo, Delta sound (click) and wood sounds with the HRTF, for a distance of 0.5 meters and direction of 0°, using the Discrete Fourier Transform method. The x-axis represents the time in seconds (the bongo sound lasts 0.104 s, the delta and wood sounds last 0.078 s), while the y-axis represents the wave amplitude after convolution and normalization (between -1 and 1 values). Each of the three sounds is composed by the two channels (left and right) of the stereo sound.77

Figure 4.3 Graphic user interface, where the upper side of the interface represents the coordinates of the mouse, the results are in percent and the coordinates of the selected sound and the number of rested sounds that are to be reproduced.80

Figure 4.4 Average localization responses for ten listeners plotted as a function of material wood, bongo and Delta (click) in %, where(a) represents the results for the anechoic sounds and (b) represents the results of the sounds with reverberation. Errors bars are standard deviation.82

Figure 4.5 Localization accuracy for wood sound (anechoic and with reverberation). The x-axis represents the averaged localization data by subjects and y-axis represents the localization accuracy in %.....83

Figure 4.6 Listener X perception accuracy for anechoic sounds localization (a) in azimuth and (b) in distance. The x-axes are the trials and the y-axes represent the localization accuracy in %86

Figura 5.1 Representación esquemática del Sistema.91

Figura 5.2 Imagen real con los elementos del sistema: gafas (elemento sensor), auriculares y FPGA.....92

Figura 5.3 Imágenes del dispositivo en uso.....	94
Figura 5.4 Prueba de Laberinto	97
Figure 6.1 Measurement principle of the Time of Flight (TOF) method.....	104
Figure 6.2 Multiple Double Short-Time Integrated measurement principle.....	106
Figure 6.3 Head-Related Transfer Function measurement model	110
Figure 6.4 Laboratory experimental scenario	112
Figure 6.5 Average Absolute Walking Time (AWT) for laboratory experiment.....	113
Figure 6.6 Mean absolute walking times from Mobility Test I and mean number of hits. The ten clusters of bars show mean results for the ten subjects from the group A as well as the mean results from the four runs of the Group B. The cluster of bars from the second graph shows the average Number of Hits for both groups for the previous carried out runs	114

ÍNDICE DE TABLAS

Tabla 1.1 Principales necesidades de las personas con discapacidad visual	32
Tabla 1.2 El mercado potencial: la población con discapacidad visual en Europa, EEUU y Canadá (Fuente: Organización Mundial de la Salud)	33
Tabla 3.1 Propiedades técnicas del sensor 3D-CMOS	57
Tabla 3.2 Número de profundidad y las distancias correspondientes	64
Tabla 3.3 Características técnicas del banco de sonidos.....	65
Table 4.1 Time duration of the sounds.....	76
Table 4.2 Mean and minimum values in percentage for all three rounds for anechoic and reverberant situations.....	83
Table 4.3 Mean localization performances in percentage and standard deviation for wood, bongo and Delta material sound in anechoic environment for listener X.....	85
Table 6.1 Mobility test results for twenty subjects representing the Absolute Walking TIME (AWT), the Number of Hits (NH) and the Number of Corrections (NOC)	116

Dedicatorias

Esta tesis se la dedico de forma muy especial a mi mujer, a mis hijos y a mis padres.

Valencia, 13 marzo de 2014

Agradecimientos

Mi más sincero agradecimiento a todas aquellas personas que, en el desarrollo de la presente tesis doctoral, me han prestado su apoyo y ayuda, muy especialmente:

En primer lugar a mi director de Tesis Dr. D. Guillermo Peris Fajarnés, por ofrecerme la posibilidad de llevar a cabo este proyecto de tesis, por su paciencia y apoyo a lo largo de todo el camino, animándome a culminar este trabajo.

A mi codirectora, Dra. Dña. Larisa Dunai Dunai, por su gran ayuda y apoyo en la elaboración de este trabajo a lo largo de estos años. También por los conocimientos y el entusiasmo en la investigación transmitidos.

También agradezco a todos los compañeros del Centro de Investigación en Tecnologías Gráficas su apoyo en el día a día.

Valencia, 13 marzo de 2014

Resumen

Uno de los sentidos más importantes para el ser humano es la vista. La falta de visión hace que el resto de los sentidos se agudicen para intentar suplir esta carencia. Las personas invidentes tienen muy agudizado el sentido del oído lo que les permite desenvolverse en el entorno con mayor destreza auditiva que las personas con visión normal y que no se han entrenado para mejorarlo.

En la presente tesis se detalla el trabajo de investigación llevado a cabo con el propósito de desarrollar un sistema de ayuda a las personas invidentes para la localización de obstáculos en su entorno habitual. Para ello, el sistema transforma en sonidos virtuales los objetos que encuentran en su camino las personas que no pueden ver.

Para alcanzar este objetivo, se entrena a la persona invidente para que se habitúe al uso de un dispositivo que transforma un entorno real en un entorno acústico-virtual. De esta manera, es capaz de detectar los objetos que le rodean y así poder deambular de forma segura por ese entorno real.

Dado que se necesitaba tener un sistema de entrada de información lo más preciso posible, se optó por un sistema láser 3D-CMOS (Three Dimensional Complementary Metal Oxide Semiconductor), compuesto por 1x64 pixeles del láser miniaturizado e integrado en unas gafas. La información recibida por los láseres se transmite a una memoria Correlated Double Sampling (CDS), en la que la distancia entre el sistema y los objetos se calcula mediante el método Tiempo de Vuelo “Time-of-Flight” (ToF). Este tipo de dispositivo se caracteriza por su alta precisión en el cálculo de distancias y permite llegar a obtener errores inferiores al 1% en el dispositivo utilizado para esta tesis y desarrollado por SIEMENS.

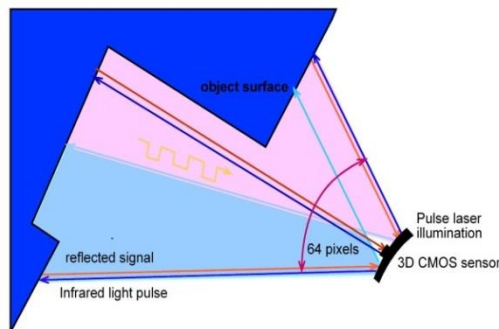


Figura 1 Método de medición por Tiempo de Vuelo (ToF).

Una vez obtenidos los parámetros de entrada, se necesitaba transformar en sonidos virtuales la ubicación exacta de los objetos en distancia y ángulo. Esto se consiguió mediante las Funciones de Transferencia (HRTF) medidas en una sala anecoica con la ayuda de un robot KEMAR. Estas funciones consiguen transformar en sonidos las posiciones captadas con el dispositivo de entrada ubicándolos en su posición real.

Los resultados presentados en esta tesis demuestran que es posible entrenar a una persona con discapacidad visual para que pueda utilizar este sistema como una herramienta fiable para la localización de obstáculos en su entorno habitual.

Resum

Un dels sentits més importants per al ser humà és la vista. La falta de visió fa que la resta dels sentits s'aguditzen per a intentar suplir esta carència. Les persones invidents tenen molt aguditzat el sentit de l'oïda, cosa que els permet desembolicar-se en l'entorn amb major destresa auditiva que les persones amb visió normal i que no s'han entrenat per a millorar-ho.

En la present tesi es detalla el treball d'investigació dut a terme amb el propòsit de desenrotllar un sistema d'ajuda a les persones invidents per a la localització d'obstacles en el seu entorn habitual. Per a això, el sistema transforma en sons virtuals els objectes que troben en el seu camí les persones que no poden veure.

Per a aconseguir este objectiu, s'entrena a la persona invident perquè s'habitue a l'ús d'un dispositiu que transforma un entorn real en un entorn acusticovirtual. D'esta manera, és capaç de detectar els objectes que li envolten i així poder deambular de forma segura per eixe entorn real.

Atés que es necessitava tindre un sistema d'entrada d'informació com més precís millor, es va optar per un sistema làser 3D-CMOS (Three Dimensional Complementary Metall Oxide Semiconductor), compostat per 1x64 píxels del làser miniaturitzat i integrat en unes ulleres. La informació rebuda pels làsers es transmet a una memòria Correlated Double Sampling (CDS), en la qual la distància entre el sistema i els objectes es calcula per mitjà del mètode Temps de Vol "Time-of-Flight" (ToF). Este tipus de dispositiu es caracteritza per la seua alta precisió en el càlcul de distàncies i permet arribar a obtindre errades inferiors al 1% en el dispositiu utilitzat per a esta tesi i desenrotllat per SIEMENS.

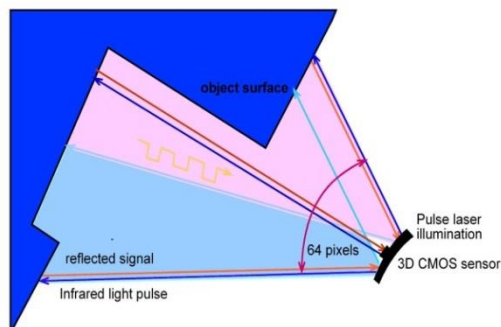


Figura 1 Mètode de mesurament per Temps de Vol (ToF).

Una vegada obtinguts els paràmetres d'entrada, calia transformar en sons virtuals la ubicació exacta dels objectes en distància i angle. Açò es va aconseguir per mitjà de les Funcions de Transferència (HRTF) mesurades en una sala anecoica amb l'ajuda d'un robot KEMAR. Estes funcions aconseguixen transformar en sons les posicions captades amb el dispositiu d'entrada ubicant-los en la seua posició real. Els resultats presentats en esta tesi demostren que és possible entrenar una persona amb discapacitat visual perquè pugui utilitzar este sistema com una ferramenta fiable per a la localització d'obstacles en el seu entorn habitual

Abstract

The sight is one of the most important senses for humans. The lack of sight makes the other senses to be heightened to try to fill this gap. Blind people have very acute sense of hearing which enables them to deal with the environment having more listening skills that people with normal vision and not trained to improve it.

This thesis research work has been carried out in order to develop a system to aid the blind people to the location of obstacles in their usual environment. For this, the system transforms into virtual sounds the objects found in their path by people that cannot see.

To achieve this goal, it trains the blind person to get used to using a device that transforms a real environment in an acoustic-virtual environment. Thus, it is able to detect objects around you so you can wander safely through the real environment.

Since it needed a data entry system as accurate as possible, we chose a 3D-CMOS (Three Dimensional Complementary Metal Oxide Semiconductor) laser system consisting of 1x64 pixels of miniaturized laser and we integrated it into glasses. The information received by the laser is transmitted to a memory Correlated Double Sampling (CDS), in which the distance between the system and the object is calculated using the Time-of-Flight (TOF) method. This type of device is characterized by its high accuracy in the calculation of distances, providing access to obtain less than 1% errors in the device used for this thesis and developed by SIEMENS.

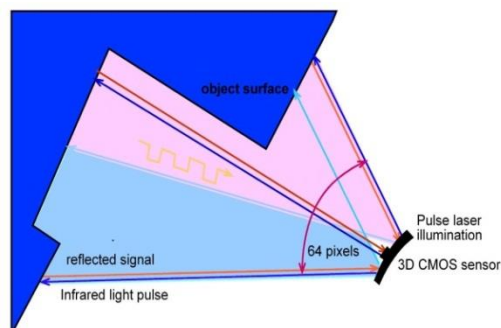


Figure 1 Measurement principle of the Time of Flight (ToF) method.

After obtaining the input parameters, it was needed to transform into virtual sounds the exact location of objects in distance and angle. This was achieved through the Transfer Functions (HRTF) measured in an anechoic room with the aid of a KEMAR robot. These functions get converted into sounds positions taken with the input device by placing them in their actual position. The results presented in this thesis show that it is possible to train a person with visual disabilities to be able to use this system as a reliable tool for locating obstacles in their usual environment.

1. INTRODUCCIÓN

En el mundo existen 285 millones de personas con discapacidad visual. De ellas, 45 millones de personas son invidentes lo que representa casi el 16% del total de personas con discapacidad visual [1].

La principal motivación que me ha llevado a la realización de esta tesis ha sido poder llegar a mejorar la calidad de vida de las personas invidentes. En este capítulo se describe con detalle la cantidad de personas invidentes y o con algún tipo de discapacidad visual, como la baja visión, que hay en el mundo. De hecho, se calcula que en la próxima década, debido al aumento de la esperanza de vida en los países desarrollados, esta deficiencia va a incrementarse. Enfermedades como la degeneración macular asociada a la edad (DMAE), la retinosis pigmentaria, la miopía magna, la tensión ocular, etc... explican este incremento en el número de discapacitados visuales. Prueba de ello es que la Organización Mundial de la Salud (OMS) ha implantado el proyecto Visión 2020 del Organismo Internacional de Prevención de la Ceguera [2], que ofrece a los gobiernos de todo el mundo un conjunto de instrumentos para ayudar a prevenir y tratar millones de casos de ceguera en previsión del gran aumento que se está detectando en esta deficiencia.

Esta tesis se enmarca como investigación dentro del proyecto europeo CASBlIP (Sistema de Ayuda Cognitiva para Ciegos), dentro de 6º Programa Marco de la Unión Europea. El objetivo principal del proyecto CASBlIP es desarrollar un sistema capaz de interpretar y gestionar la información del mundo real a partir de diferentes fuentes para apoyar la movilidad ayuda a toda la comunidad con discapacidad visual.

El sistema está compuesto por un lado de un dispositivo de visión estereoscópica para capturar la información del entorno, con procesamiento de imágenes para cálculo de distancias en campo lejano, y por otro lado de un sensor CMOS 3D basado en tecnología láser para adquirir información de distancia en campo cercano. Por otra parte, consta de un sistema de ubicación con tecnología de Posicionamiento Global. El interfaz de salida está compuesto por un interfaz de audio para transformar el entorno real capturado en sonidos.

Esta Tesis está centrada en el dispositivo 3D CMOS como dispositivo de entrada y la transformación de la información recibida un “mapa sonoro virtual” del entorno próximo, en un radio de entre 0,5 y 5 metros.

Como parte de la introducción, se describe de forma detallada la tecnología utilizada, haciendo especial hincapié en el sistema 3D-CMOS usado como dispositivo de captura del entorno y de la formulación acústica HRTF utilizada para la transformación del entorno en un espacio acústico-virtual.

Por último se detalla el listado de las publicaciones realizadas durante el desarrollo de la tesis.

1.1 Generalidades

La discapacidad visual hace referencia a la carencia, disminución o defecto de la visión de forma que según el grado de pérdida de la visión, se distinguen 2 colectivos con necesidades y características distintas:

- Las personas invidentes o ciegas (con pérdida total de visión).
- Las personas con privación parcial de la vista, lo que se denomina baja visión (que puede ser grave o moderada).

En el mundo existen 285 millones de personas con discapacidad visual, de las cuales 246 millones presentan baja visión y 39 millones de personas son invidentes lo que supone que el 0,7 % de la población mundial son invidentes [1]. Según los últimos datos publicados por la Organización Mundial de la Salud, solamente, en Estados Unidos, Canadá y en Europa hay más de 2 millones de ciegos y 29 millones de personas con una discapacidad visual [1-2].

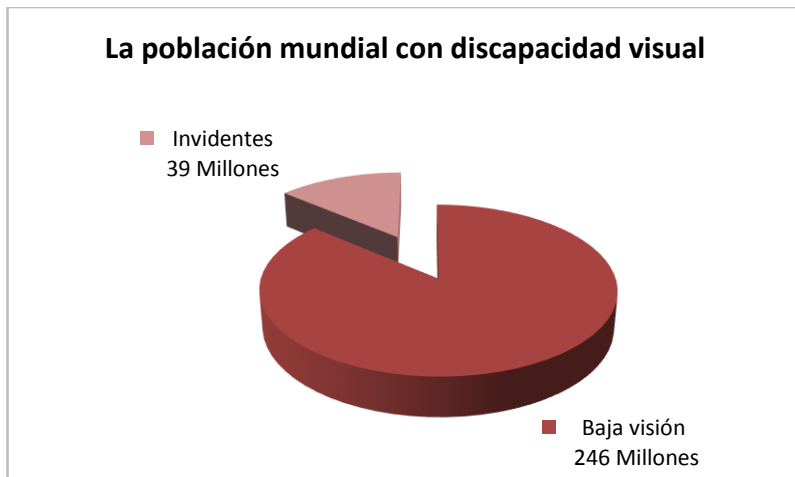


Figura 1.1 La población mundial con discapacidad visual. (Fuente: Organización Mundial de la Salud) [1-2].

El número de personas ciegas y con discapacidad será un 30% mayor en el año 2020 [2], entre otros factores, motivado por la mayor longevidad de la población. Solamente en EEUU el crecimiento de la población mayor de 65 años en los últimos años ha pasado del 13% al 19%, cuando en China el salto ha sido del 7% al 13%.

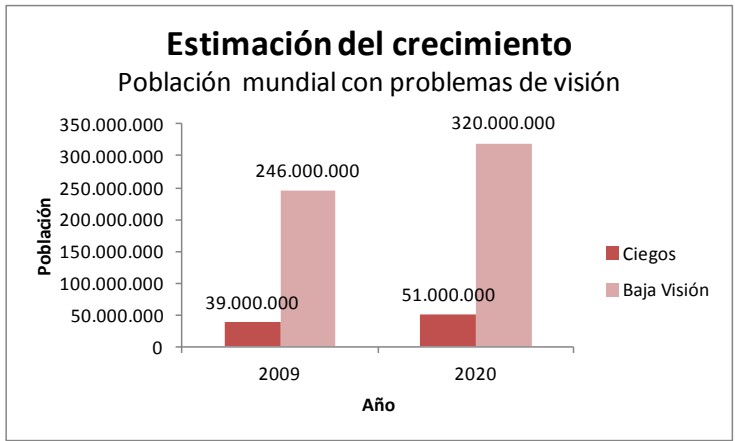


Figura 1.2 Estimación del crecimiento de la población mundial con problemas de visión (Fuente: Organización Mundial de la Salud, World Blind Union) [1,3].

1.1.1 Las características de la discapacidad visual

La pérdida de la visión afecta directamente a la vida social de las personas. En estos casos, por la falta de la información visual, aparecen problemas de lectura, escritura, reconocimiento de personas o lugares y movilidad. En cuanto a las causas de la pérdida de la visión, las más significativas son las siguientes:

En cuanto a las principales causas de la ceguera los más comunes son:

- Las cataratas suponen casi el 40% de las causas y suelen aparecer con la edad. Pueden ser corregidas mediante intervención quirúrgica.

- Los errores de refracción, miopía, hipermetropía, astigmatismo representan un 18%.
- Glaucoma un 10%.
- La degeneración macular asociada a la edad (DMAE), 7%.
- Retinopatía diabética (RD) 4%
- Opacidad de la Córnea 4%
- Tracoma (Enfermedad ocular infecciosa causada por la bacteria Chlamydia trachomati) 3%
- Ceguera Infantil 3%
- Otros 11%

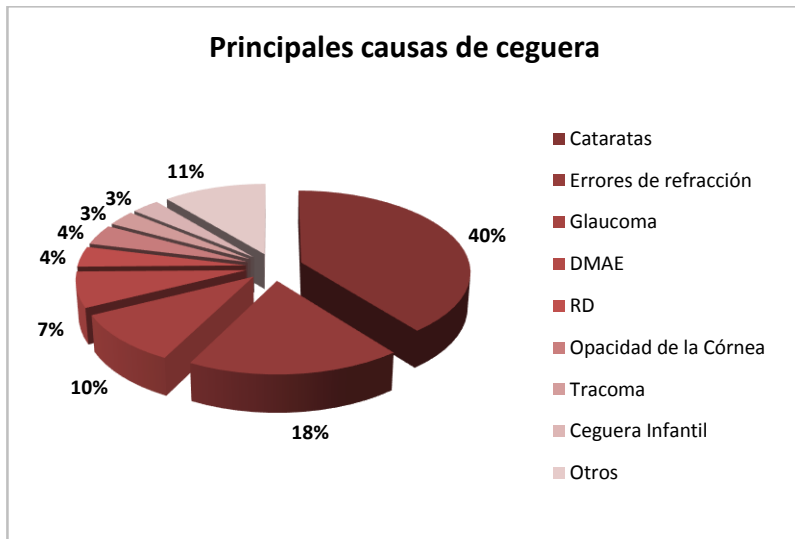


Figura 1.3 Principales causas de ceguera [5].

En cuanto a los principales problemas que generan la baja visión, destacan:

- La visión borrosa, que representa la pérdida de la agudeza visual y la incapacidad para visualizar pequeños detalles.
- Los puntos ciegos, que se caracterizan por los agujeros oscuros en el campo de visión que impiden ver algo en este campo.
- Los halos.
- Las moscas volantes, que representan pequeñas estrellitas que flotan a través del ojo. Este síntoma puede llevar al desprendimiento de la retina que crea unos destellos de luz o una sensación de sombra en una de las partes del campo visual.

La deficiencia visual de estas personas supone la ausencia completa o casi completa de visión, es decir, ceguera o visión insuficiente (también llamada baja visión), debido a múltiples problemas como la visión borrosa, los agujeros oscuros en el campo de visión, las pequeñas estrellas que flotan en el ojo, los destellos de luz o la sensación de sombra en una de las partes del campo visual.

A pesar del apoyo que prestan las sociedades más desarrolladas para solucionar los problemas de visión, estas personas padecen una serie de incomodidades para comunicarse, tener una vida social, reconocer personas y lugares, leer o escribir, incluso crear una familia y, en el caso de las personas invidentes, tener una vida independiente debido a las dificultades de movilidad y de autonomía.

Diversos estudios sobre la ceguera y la discapacidad visual grave realizados a nivel mundial, entre los que destacan los realizados por la Universidad de Nottingham en Reino Unido, revelan el bajo nivel de autonomía de las personas invidentes y lo poco que ha mejorado su movilidad desde hace 40 años debido a que los productos disponibles en el mercado como los bastones para ciegos o los perros guía son insuficientes o solucionan parcialmente sus problemas de movilidad.

Los bastones para ciegos, a pesar de tener un precio competitivo, tienen importantes limitaciones en cuanto a su funcionalidad y alcance como por

ejemplo que la información percibida sea como máximo de 1,5 metros de distancia.

En cuanto a los perros guía, además de tener limitaciones funcionales como la capacidad de distinguir los colores rojo y verde de los semáforos, tienen un coste de 30.000 euros por unidad entrenada, lo que limita su oferta y no cubre el total de la demanda existente. Por poner un ejemplo, España tiene una capacidad de entrenamiento de perros guía de 120 unidades al año, lo que supone que solamente 1.000 personas dispongan de perros guía como auxiliar de movilidad. Una oferta muy por debajo de las 70.000 personas invidentes y las 200.000 que tienen una grave discapacidad visual en nuestro país [4-5].

También se constata que las personas con discapacidad visual grave no utilizan todas las herramientas existentes para compensar su déficit. En ocasiones, el resto de capacidad visual que tienen disponible no representa una ventaja sino todo lo contrario, es decir, no ven lo suficiente para manejarse como vidente pero no manejan los instrumentos de los que podría beneficiarse una persona ciega rehabilitada.

Esta situación muestra la existencia de una importante demanda de dispositivos y productos de ayuda para este colectivo que el mercado no ha cubierto. Una necesidad que en los países europeos, EEUU y Canadá supondría un importante mercado potencial debido a su capacidad económica y a la sensibilidad social hacia estas personas. Esta situación ha generado e impulsado desde hace años el desarrollo de una tecnología que mediante la transformación de las imágenes tridimensionales en sonidos, permite el uso de la capacidad natural del ser humano de identificar la posición de objetos en el espacio.

Sin embargo, la aplicación de esta tecnología no ha estado exenta de problemas y dificultades técnicas que han impedido su total desarrollo y su posterior comercialización destacando los siguientes motivos:

- El elevado número de ideas y prototipos que utilizan esta tecnología no han terminado de desarrollarse y, por lo tanto, no han permitido su comercialización al no disponer de un respaldo industrial y empresarial. Es el caso del dispositivo láser Laser Cane, los dispositivos de navegación MoBIC y Personal Guidance System basados en la tecnología GPS; los dispositivos basados en el sistema de estereovisión vOICE y EPEL o los dispositivos

SWAN y NAVY que combinan varios sistemas de lectura y representación del entorno.

1.1.2 Necesidades de las personas con discapacidad visual

Los objetivos más importantes en la vida cotidiana de las personas ciegas o con discapacidad visual grave son según los estudios realizados por los organismos, organizaciones o fundaciones de ciegos (Fundación Once [6] y la Royal National Institute of Blind People [7]):

Tabla 1.1 Principales necesidades de las personas con discapacidad visual.

Principales necesidades
1. Tener un vida independiente y productiva
2. Comunicar
3. Crear una familia
4. Tener una vida social
5. Movilidad
6. Leer
7. etc.

Para conseguir estos objetivos, los discapacitados visuales necesitan mucho esfuerzo y ayuda tanto psicológica como física. El problema más importante para conseguir estos objetivos y necesidades es la movilidad.

A continuación se muestra una tabla con el número total de personas con discapacidad visual por país detallando las personas con baja visión y con ceguera.

Tabla 1.2 La población con discapacidad visual en Europa, EEUU y Canadá (Fuente: Organización Mundial de la Salud) [1].

POBLACIÓN CON DISCAPACIDAD VISUAL				
CONTINENTE	PAÍS	Ciegos	Baja visión	TOTAL discap visual
Europa	España	70.000	1.616.259	1.686.259
	Italia	144.112	2.087.381	2.231.493
	Francia	156.695	2.269.640	2.426.335
	Reino Unido	144.824	2.097.698	2.242.522
	Alemania	150.000	2.819.856	2.969.856
	Resto Europa	434.473	6.293.094	6.727.567
	Subtotal	1.100.104	17.183.928	18.284.032
América	EEUU	819.174	10.615.727	11.434.901
	Canada	80.722	1.169.209	1.249.931
Total		2.000.000	28.968.864	30.968.864

1.1.3 Dispositivos tecnológicos para las personas invidentes

La demanda potencial de dispositivos tecnológicos para la vista es desconocida al no estar desarrollado ni comercializado ningún producto en el mercado para personas ciegas.

Los productos que se han desarrollado hasta el momento no han llegado más allá de la fase de prototipo e investigación.

Desde los años 50, se han desarrollado más de 45 dispositivos de navegación basados en las siguientes tecnologías:

- Sensores.
- Ultrasonido o laser: Laser Cane.
- Visión artificial.
- Mapas electrónicos.
- GPS: MoBIC y el Personal Guidance System.
- Sistemas de estereovisión: VOICE y EPEL.
- Sistemas de lectura y representación del entorno: SWAN y NAVY.

La falta de éxito de estos prototipos se debe a varios motivos:

- La existencia de limitaciones tecnológicas. Los sistemas de representación de la información recibida del entorno son muy confusos y requieren del usuario una alta atención de la información recibida por el dispositivo que le impide detectar la información del entorno, el tráfico o las señales de tráfico entre otros.
- La representación mediante voz sintética es larga y requiere mucho tiempo de reproducción.
- Los dispositivos táctiles son grandes e incómodos.
- Los dispositivos que tienen un alto funcionamiento para el entorno abierto no son fiables para entornos cerrados.

Todos estos dispositivos están en fase de desarrollo por lo que además tienen un coste muy elevado.

1.1.4 Productos complementarios a la navegación

Los perros guía

En España, la demanda de perros guía es mucho más alta que la oferta. De hecho se entregan únicamente unos 120 al año y en este momento la cifra de usuarios con perro guía supera las 1.000 personas. Para que un perro guía pueda ser donado a un invidente, se exige que éste pueda manejar un bastón y se encuentre en perfecto estado físico. El que un ciego reciba un perro guía es un proceso que suele demorarse varios años, ya que el número de perros guía está en dependencia directa de los perros que nazcan en cada camada, su crianza y su educación posterior.

La compra de un perro guía como su entrenamiento tienen un coste de alrededor de 30.000 euros por unidad. El sistema que se analiza en esta tesis se va a convertir, sin duda, en una ayuda potencialmente superior a la que puede aportarles un perro guía.

El bastón

El bastón supone para el ciego una prolongación de su cuerpo. Esta herramienta táctil, fabricada con fibras sintéticas, guía los pasos del invidente,

haciéndole notar sobre qué tipo de superficie está caminando y proporcionándole asimismo información acerca de su ubicación. El contacto del puntero con el suelo remite unas vibraciones al ciego, que son más intensas cuanto más dura es la superficie sobre la que se encuentra.

En el mercado, existen actualmente distintos modelos de bastones, cada uno de los cuales presenta un tipo diferente de agarre. Su diseño es uno u otro en función del servicio que están destinados a proporcionar, dependiendo del grado de ceguera del usuario. Así, hay un modelo denominado B1, que está diseñado para guiar a ciegos totales, un modelo B2 dirigido a aquéllos que distinguen levemente luces y sombras, y finalmente un modelo B3 que orienta los pasos de los invidentes que todavía presentan algún resto de visión. El precio de los bastones supera los 30 euros.

1.2 HRTF

Los seres humanos percibimos el sonido mediante nuestro sistema auditivo como variaciones de presión del aire causadas por las vibraciones de un objeto. La percepción de los sonidos comienza cuando la forma de onda llega al oído, variando su presión.

Con el fin de localizar una fuente de sonido, el sistema auditivo humano se basa principalmente en los siguientes aspectos [8]:

- **Diferencia Tiempo Interaural (ITD):** La diferencia de tiempo entre la llegada del sonido a cada uno de los oídos.
- **Diferencia de Nivel Interaural (ILD):** La diferencia de nivel de presión sonora (SPL) entre los sonidos en ambos oídos.
- **Interacción con la visión:** Podemos determinar la ubicación de una fuente de sonido que podemos ver.
- **Head Related Transfer Functions (HRTF):** la compleja interacción de una onda de sonido con el torso, los hombros, la cabeza y todo el pabellón auricular (oído externo) de un oyente, esencialmente, el pabellón de la oreja. Cada oreja filtra cada onda de sonido que pasa a través de él de alguna manera única a la posición de fuente de sonido. Teniendo en cuenta estas señales filtradas, el cerebro calcula la posición exacta en 3D de una fuente de sonido con respecto al oyente.
- **Reverberación:** Reflexiones de las ondas de sonido de otros objetos en el ambiente (por ejemplo, las paredes de una habitación).

Otros elementos que afectan a localización de sonido de acuerdo con la descripción hecha [9] son la cabeza, así como el movimiento de la fuente, el efecto precedente y percepción de la distancia auditiva, con el fin de establecer una localización completa de fuentes de sonido por el sistema auditivo humano [10].

Con la finalidad de obtener unos sonidos virtuales situados en la posición real de un objeto, se han utilizado las medidas de las Funciones de Transferencia (HRTF) en una sala anecoica con la ayuda de un robot KEMAR. Un Maximum Length Binary Sequence (MLBS) se utilizó como fuente sonora para la medida de los HRTFs.



Figura 1.4 La cámara anecoica con bajo nivel de reverberación y sistema robótico. Las paredes y el suelo de la sala están cubiertas con cuñas que absorben el sonido. El sistema robótico conjunto puede ser movido para colocar el altavoz en cualquier ubicación en radio de 5 metros. Una silla para el maniquí se representa en una plataforma fija.

Con respecto a la formulación de la onda se procede de la siguiente forma; sea $x_1(t)$ la señal reproducida por el altavoz e $y_1(t)$ la señal grabada por el micrófono colocado en el canal audio del maniquí robot. Se diseña un filtro digital para transformar $x_1(t)$ en la señal $x_2(t)$ reproducida por el auricular. Sea X_1 la representación del $x_1(t)$ en el dominio de la frecuencia, e Y_2 la representación de $x_2(t)$. Al reproducir $x_1(t)$, la respuesta que se detecta se puede representar como:

$$Y_1 = X_1 LFM \tag{1.1}$$

Donde L representa la Función de Transferencia del altavoz con toda la cadena de audio de reproducción, F es la Función de Transferencia del espacio libre entre el altavoz y el canal auditivo, y M es la Función de Transferencia del micrófono y toda la cadena de audio de grabación. De la misma forma se calcula la respuesta grabada por el micrófono al reproductor $x_2(t)$.

$$Y_2 = X_2HM. \quad (1.2)$$

Donde H representa la Función de Transferencia del auricular al canal auditivo. Si $Y_1=Y_2$ obtenemos el filtro digital:

$$T = \frac{LF}{H}. \quad (1.3)$$

Finalmente, al someter el sistema a una Secuencia Binaria de Máxima Longitud (MLBS) que denominamos $s(n)$, la respuesta del impulso periódica $h(n)$ será:

$$h(n) = \Omega_{sy}(n) = s(n) * \Phi y(n). \quad (1.4)$$

Donde Φ es la correlación cruzada circular periódica. Para reducir el tiempo de cálculo se utiliza la Transformada Hadamard Rápida (FHT).

$$h(n) = \frac{1}{(L+1)s[0]} P_2 \langle S_2 \{ H_{L+1} [S_1 (P_1 y(n))] \} \rangle. \quad (1.5)$$

Donde P son las matrices de permutación y las S -las matrices de redimensionamiento, H_{L+1} representa la matriz Hadamard de orden $L+1$. Para realizar la convolución con el sonido se ha utilizado la Transformada Fourier Discreta [4].

1.3 3D-CMOS – Time of Flight

1.3.1 3D-CMOS

El dispositivo utilizado como dispositivo de entrada es un sistema láser 3D-CMOS (Three Dimensional Complementary Metal Oxide Semiconductor), compuesto por 1x64 pixeles del láser miniaturizado e integrado en unas gafas en un plano horizontal cubriendo una distancia máxima de 5 metros y una cobertura en azimut de 64°, como se puede ver en el dispositivo final de muestra en la Figura 1.5. La información recibida por los láseres se transmite a una memoria Correlated Double Sampling (CDS), en la que la distancia entre el sistema y los objetos se calcula mediante el método Time-of-Flight (ToF). Este tipo de dispositivo se caracteriza por su alta precisión en el cálculo distancias llegando a obtener errores en el dispositivo utilizado para esta tesis y desarrollado por SIEMENS inferiores al 1%.



Figura 1.5 Imagen real con los elementos del sistema: gafas (elemento sensor 1x64 pixeles del láser miniaturizado e integrado).

El sensor 3D-CMOS ha sido desarrollado por SIEMENS dentro del proyecto CASBlIP. El aspecto innovador de este dispositivo es su tamaño y su uso para la navegación para las personas ciegas. Una de las principales ventajas es

su precisión en la detección, (el láser tiene una precisión de 99%) y el procesamiento es en tiempo real.

El principio de medición se basa en el tiempo de vuelo, medición de la luz láser de pulso modulada con un sensor CMOS fotosensible y de alta velocidad de

infrarrojos láser de pulso de iluminación. Los impulsos consisten en una luz infrarroja emitida por el sensor CMOS. Cuando cualquier objeto aparece en la parte delantera del láser, los impulsos infrarrojos chocan con el objeto y envían los impulsos infrarrojos de nuevo al sistema. Cuando los impulsos llegan al sistema, se calcula la distancia entre el dispositivo y el objeto. Las señales analógicas de varios pulsos láser se promedian en un chip para reducir la potencia del láser requerida y también para aumentar la precisión de la medición. La ventaja del uso de estos sensores es proporcionar una distancia exacta tanto en el plano vertical como en el plano frontal. Además, se reduce el tiempo de procesamiento necesario para el cálculo.

Algunas de las aplicaciones en las que se está utilizando esta tecnología es en la detección de colisiones de automóviles y detección de peatones para la prevención de atropello de forma automática, como se observa en el ejemplo Figura 1.6.



Figura 1.6 Coche de prueba con una cámara PMD para seguridad de los peatones y la detección de Pre-colisión, respectivamente [12].

1.3.2 Time of Flight

El cálculo del tiempo de vuelo que se utiliza para calcular la distancia a las que se encuentran los objetos lo podemos aproximar a la interferometría. La interferometría es una técnica utilizada en astronomía que combina la luz de diferentes receptores, como pueden ser los telescopios y radiotelescopios, para obtener imágenes de mayor resolución.

Con esta técnica podemos medir una distancia absoluta si logramos medir el tiempo absoluto que un pulso de luz necesita para viajar a un objeto y vuelta de nuevo a un receptor. Esta medición indirecta de la distancia es posible, ya que sabemos la velocidad de la luz con gran precisión:

$$c = 3 \cdot 10^8 \text{ m/s} = 2 \cdot 150 \text{ m/ms} = 2 \cdot 0,15 \text{ m/ns} = 2 \cdot 0,15 \text{ mm/ps.} \quad [13-14]$$

Si la fuente de luz activa y el receptor se encuentran muy cerca el uno al otro, esto facilita una configuración compacta y evita efectos de sombreado. Se ilustra el principio básico de un sistema de Time of Flight. Una fuente emite un pulso de luz y comienza un cronómetro de alta precisión en el detector. El pulso de luz viaja al objeto y vuelve. La recepción de la luz por el mecanismo detector detiene la medición del tiempo y el cronómetro ahora muestra el tiempo de vuelo. Teniendo en cuenta el hecho de que la luz viaja por el camino dos veces (adelante y atrás) un tiempo medido de 6,67 ns corresponde a una distancia de 1 m y se requiere una precisión de picosegundos para conseguir resoluciones de 1 mm. Una característica esencial de esta configuración es el hecho de que el emisor y el detector operan sincrónicamente [15].

1.4 Publicaciones científicas y registro propiedad intelectual

En este punto se detallan los artículos y congresos, seguidos de registro de propiedad intelectual, publicados durante la realización de la presente tesis.

ARTÍCULOS DE REVISTA JCR

Ismael Lengua, Larisa Dunai, Guillermo Peris-Fajarnés, Beatriz Defez.
**DISPOSITIVO DE NAVEGACIÓN PARA PERSONAS INVIDENTES
BASADO EN LA TECNOLOGÍA TIME OF FLIGHT.**
Dyna, año 80, Edición 179, pp. 33-41, Medellín, junio, 2013. ISSN 0012-73.
ÍNDICE DE IMPACTO (JCR 2012): 0.175
ENGINEERING, MULTIDISCIPLINARY: 78/83

Larisa Dunai, Guillermo Peris-Fajarnes, **Ismael Lengua**, Ignacio Tortajada Montañana.
**MATERIAL SOUND SOURCE LOCALIZATION THROUGH
HEADPHONES.**
Acoustical Physics, 2012, Vol. 58, No. 5, pp. 610–617. ISSN 1063-7710.
ÍNDICE DE IMPACTO (JCR 2012): 0.421
ACOUSTICS: 25/30

ARTÍCULOS INDEXADOS

Ismael Lengua, Larisa Dunai, Guillermo Peris-Fajarnés, Fernando Brusola.
**LOCALIZACIÓN DE SONIDOS EN ENTORNO ABIERTO POR PERSONAS
CIEGAS.**
Revista de Acústica de la Sociedad Española de Acústica. Vol. 45. Núm 3-4 2013.
I.S.S.N. 0210-3680.
LATINDEX

ARTÍCULOS DE CONGRESOS INTERNACIONALES Y NACIONALES

Larisa Dunai, Beatriz Defez, **Ismael Lengua**, Guillermo Peris-Fajarnes.

3D CMOS SENSOR BASED ACOUSTIC OBJECT DETECTION AND NAVIGATION SYSTEM FOR BLIND PEOPLE.

IECON 2012 - 38th Annual Conference on IEEE Industrial Electronics Society. 4208-4215 Montreal, QC Date of Conference: 25-28 Oct. 2012, ISBN: 978-1-4673-2419-9.

Scopus Coverage Years: from 1989 to 1991, from 1993 to 2007, from 2009 to 2012

SJR (SCImago Journal Rankings) 2012: 0.279

SNIP (Source Normalized Impact per Paper) 2012: 0.210

Larisa Dunai, **Lengua Ismael**, Guillermo Peris-Fajarnes, Beatriz Defez

OBJECT DETECTION AND REPRESENTATION PROCEDURE FOR A NAVIGATION DEVICE FOR BLIND PEOPLE

Sociedad Española de Acústica (SEA)

44º Congreso Español de Acústica. Encuentro Ibérico de Acústica. EAA European Symposium on Environmental Acoustics and Noise Mapping. pp. 1131 – 1140. 2013.

ISSN 978-84-87985-23-2.

Larisa Dunai, **Lengua Ismael**, Guillermo Peris-Fajarnes, Beatriz Defez

VIRTUAL SOUND LOCALIZATION BY BLIND PEOPLE

EDITORIAL: Socieda de Portuguesa de Acústica (SPA)

VIII Congreso Ibero-americano de Acústica (Acústica 2012) pp. 41 – 41. 2012.

ISSN 978-989-20-3284-9.

Larisa Dunai, Guillermo Peris-Fajarnés, **Ismael Lengua**, Fernando Brusola.

EYE 21 DISPOSITIVO PARA PERCIBIR EL ENTORNO MEDIANTE SONIDOS ACÚSTICOS

I Congreso Internacional Universidad y Discapacidad (CIUD) pp.1-1. 2012.

REGISTRO DE PROPIEDAD INTELECTUAL

Coautor del software “EYE21” para la ayuda a la movilidad de personas ciegas mediante el uso de cámaras y sonido espacializado, registrado en el Centro de Transferencia Tecnológica (CTT) de la Universidad Politécnica de Valencia con registro número R-15882-2011.

1.5 Referencias

- [1] Organización Mundial de la Salud. Datos sobre la ceguera y discapacidad visual. <http://www.who.in>. [Consulta: jueves, 05 de septiembre de 2013]
- [2] VISION 2020: The right to sight. Organización Mundial de la Salud. Datos sobre la ceguera y discapacidad visual. <http://www.iapb.org/vision-2020>. [Consulta: jueves, 05 de septiembre de 2013]
- [3] World Blind Union. Datos sobre la ceguera y discapacidad visual. <http://www.worldblindunion.org/>. [Consulta: jueves, 09 de septiembre de 2013]
- [4] Instituto Nacional de Estadística. Datos sobre la ceguera y discapacidad visual. <http://www.ine.es/>. [Consulta: jueves, 09 de septiembre de 2013]
- [5] Fundación Retinaplus+, Ernst & Young (Centro de investigación, Ciencias de la Salud). Informe de la ceguera en España. Octubre 2012. SP1205-42E. ED 0814.
- [6] Fundación ONCE. Objetivos más importantes en la vida cotidiana de las personas ciegas. <http://www.fundaciononce.es/>. [Consulta: jueves, 05 de septiembre de 2013]
- [7] Royal National Institute of Blind People. Objetivos más importantes en la vida cotidiana de las personas ciegas. <http://www.rnib.org.uk/>. [Consulta: jueves, 05 de septiembre de 2013]
- [8] Bill Kapralos, Michael R.M. Jenkin,. “AUDITORY PERCEPTION AND SPATIAL (3D) AUDITORY SYSTEMS”. 2003.
- [9] R. Begault. “3D SOUND FOR VIRTUAL REALITY AND MULTIMEDIA”. Durand National Aeronautics and Space Administration, 2000. NASA/TM.
- [10] Dea L. Dunai “EXPERIMENTAL LISTENING TESTs OF HEARING MATERIAL IN ELECTRONIC TRAVEL AID SYSTEMS”. 2008.

- [11] Ismael Lengua, Larisa Dunai, Guillermo Peris-Fajarnés, Beatriz Defez. “DISPOSITIVO DE NAVEGACIÓN PARA PERSONAS INVIDENTES BASADO EN LA TECNOLOGÍA TIME OF FLIGHT”. Dyna, año 80, Edición 179, pp. 33-41, Medellín, junio, 2013. ISSN 0012-73.
- [12] T. Ringbeck, B. Hagebecker, “A 3D TIME OF FLIGHT CAMERA FOR OBJECT DETECTION” Optical 3-D Measurement Techniques 09-12.07.2007 ETH Zürich. Plenary Session 1: Range Imaging I
- [13] [BRK] Brockhaus, “Naturwissenschaften und Technik”, Bd. 3, ISBN 3-7653-0450-6, Mannheim, (1989).
- [14] [BRR] H. Breuer, “dtv-Atlas zur Physik, Tafeln und Texte”, Bd.1, 2.Auflage, ISBN 3-423-03226-x, (1990).
- [15] R. Lange, “3D Time-of-Flight Distance Measurement with Custom Solid-State Image Sensors in CMOS/CCD-Technology” A dissertation for the degree of doctor en Technical Science, (2000)

2. OBJETIVOS

Para empezar a desarrollar la investigación de esta tesis, se planteó cuáles eran los elementos necesarios para diseñar un sistema que fuera capaz de reproducir mediante sonidos virtuales un entorno real. Con ello se pretendía que una persona invidente fuera capaz de detectar los obstáculos del entorno y poder deambular por ese espacio con total seguridad. Todo ello partió del desarrollo del proyecto europeo CASBlIP.

Para llegar a alcanzar el planteamiento inicial se planificó los siguientes hitos:

1. Seleccionar una electrónica, como dispositivo de entrada, con la finalidad de captar el entorno y obtener una respuesta de la posición y distancia a la que se encuentran los obstáculos en tiempo real.
2. Qué sonidos son los más adecuados para transformar un entorno real en un entorno acústico virtual y cómo posicionarlos.
3. Cómo entrenar a las personas invidentes en un entorno acústico virtual.
4. Cómo evaluar el funcionamiento del sistema.

Una vez definido el planteamiento inicial, para resolver estos retos se definieron los siguientes objetivos detallados a continuación:

1. Determinar si la electrónica basada en un dispositivo 3D-CMOS mediante procesamiento por Tiempo de Vuelo (ToF) es lo suficientemente rápida para considerarse en tiempo real y si la medición de distancias es lo suficientemente precisa.
2. Confirmar si el tipo de sonido que se puede llegar a utilizar como sonido virtual de referencia para la localización del entorno puede afectar a la simulación de posición de un objeto en el espacio.
3. Verificar si las Head Related Transfer Functions (HRTF) son adecuadas para el posicionamiento virtual de los objetos simulando un entorno real.
4. Realizar entrenamientos con personas invidentes en simuladores y entornos reales para localizar obstáculos.

Finalmente, determinar si el sistema planteado es fiable para que las personas invidentes puedan utilizarlo con seguridad.

3. ARTÍCULO I

Ismael Lengua, Larisa Dunai, Guillermo Peris-Fajarnés, Beatriz Defez.

DISPOSITIVO DE NAVEGACIÓN PARA PERSONAS INVIDENTES BASADO EN LA TECNOLOGÍA TIME OF FLIGHT.

Dyna, año 80, Edición 179, pp. 33-41, Medellín, junio, 2013. ISSN 0012-73.

Abstract

The paper presents a new navigation and object detector device for blind people based on Time-of-Flight technology and acoustic sounds. The device has been developed as a complementary device for blind people. Its main objective is to detect and localize the obstacles from the environment and inform to the user about their presence through acoustic sounds, both in distance and in azimuth. The device working area is from 0,5m up to 5m in distance and between 30° left and 30° right of the human head, with a precision of 0,9°.

The device architecture is based on glasses which contain a 3D-CMOS sensor, an FPGA with the whole electronics and software.

After a short experimental period it was possible to demonstrate the importance of the development of such device for the blind community. Due to the acoustic device, blind users were able to detect obstacles from the environment, localize and avoid them.

Resumen

El artículo presenta un nuevo dispositivo de navegación y detección de obstáculos para las personas ciegas, basado en la tecnología Time-of-Flight y en sonidos acústicos. El dispositivo se ha desarrollado como un dispositivo de ayuda, complementario al bastón, para las personas invidentes. Su objetivo primordial es detectar los obstáculos e informar al usuario mediante sonidos acústicos de la locación de los mismos, tanto en distancia como en dirección. El dispositivo tiene un rango de trabajo entre los 0,5 m y 5 m en distancia y entre 30° izquierda y 30° derecha en azimut, con una precisión de 0,9°. El dispositivo informa a los usuarios mediante auriculares estéreo, de la

presencia de los obstáculos situados en su camino. Está compuesto por un sistema de sensores 3D-CMOS montado en un par de gafas de sol, un FPGA que se encarga de procesar la información recibida por los sensores y transformarla en sonidos acústicos, y un par de auriculares estéreo.

Las pruebas experimentales llevadas a cabo demuestran el potencial que puede tener para las personas invidentes. Se prueba que con la ayuda del dispositivo acústico los usuarios se sienten más seguros a la hora de transitar por las calles, debido a que tienen una amplia información sobre el entorno que los rodea, más allá de la obtenida con el bastón.

3.1 Introducción

En el mundo hay más de 282 millones de personas con discapacidad visual, entre los cuales 39 millones de personas son invidentes [1]. De todos ellos, 5,9 millones viven en África, 3,2 millones en Estados Unidos y 2 millones en Europa. En Alemania hay 150 mil personas con ceguera total y 500 mil personas con discapacidad visual [2]. 5 millones de invidentes viven en China [3]. La pérdida de visión viene acompañada de una pérdida de independencia: las personas invidentes son incapaces de aprovechar las ventajas de la navegación, interactividad social y comunicación. La movilidad según Foulke [4] representa “la habilidad de viajar con seguridad, confortablemente y de forma independiente por el entorno”. Partiendo de las restricciones que conlleva la ceguera y de la definición de la movilidad, se puede deducir lo lejos que queda tal objetivo.

Con el enorme desarrollo de la tecnología desde el año 1890, se han buscado nuevas soluciones y se ha perseguido el desarrollo de nuevos Dispositivos Electrónicos de Ayuda a la Movilidad (DEAM) [5-6]. Los DEAM representan dispositivos inteligentes que tienen como objetivo superar las limitaciones humanas, percibir el entorno y representarlo a las personas invidentes mediante el tacto, vibraciones, habla o señales acústicas.

El primer DEAM, conocido como Neiszewski’s Electroftalm, se desarrolló en 1897 por Starkiewicz y Kuliszewsky. Mediante dicho dispositivo, el usuario era capaz de distinguir entre luz y oscuridad [4]. En 1912 Nye y Bliss desarrollaron un nuevo dispositivo; D’Albe’s Exploring Optophone. Este dispositivo [7], sigue una tecnología análoga al anterior, convirtiendo la luz en sonidos [8]. El principio de funcionamiento de estos dispositivos radica en el uso de un reproductor de sonidos que emite un sonido cuando disminuye la intensidad de la luz.

La mayoría de los dispositivos DEAM se clasifican según el sistema de adquisición de la información del entorno. Así, se distinguen varios grupos de DEAM: de ultrasonido, sensores o láser, Global Positioning System y de visión artificial también utilizada para los robots [9]. De los dispositivos basados en ultrasonidos los más destacados son el Lindsay Russel Pathsound [10-11], el Mowat Sonar Sensor [12] y el Sonicguide [13].

El Lindsay Russel Pathsound es un dispositivo detector de obstáculos que informa al usuario de la presencia de éstos mediante sonidos acústicos y sistema táctil. El dispositivo tiene tres tipos de sonido que representan tres distancias diferentes en dos zonas: la zona dentro del peligro y la zona fuera del peligro. La distancia máxima detectada es de 6 pies. A pesar de sus objetivos el dispositivo un ha llegado a ser comercial, ya que su diseño es muy voluminoso. El Mowat Sonar Sensor tiene una salida sonora y una táctil. El dispositivo emite un ultrasonido de forma cónica y elíptica de 15° de ancho y 30° de alto, cubriendo un área equivalente a la forma humana [13]. Su rango en distancia es de 1 a 4 metros. El dispositivo de ultrasonido más destacado es Sonicguide, también conocido como Binaural Sonar Electronic Travel Aid. El dispositivo está implementado en unas gafas y tiene como salida un sistema estereofónico. El Sonicguide emite una señal sónica de forma cónica que abarca desde 45° a la izquierda hasta 45° a la derecha con un rango de hasta cuatro metros de distancia, respecto a la línea central de visión del usuario. [14]. Según la variación del tono, el usuario es capaz de percibir y reconocer las distancias [15].

Entre los dispositivos de sensores o láser, los más destacados son los *Laser Cane*, *Talking Light*, *Pilot Light*, *SONA*; *Easy Walker*, etc.

Laser Cane o Light Amplification by Simulated emission of Radiation representa la integración del bastón con la tecnología láser [16]. El dispositivo utiliza el láser de galio-arsénico así como pequeñas baterías. El rango de detección es de hasta 4 metros e informa al usuario de la presencia de los obstáculos mediante señales acústicas [17].

Entre los dispositivos de navegación basados en Sistemas de Posicionamiento Global (GPS) destacan el *MoBIC* [18], el *GPS Braille Note* [19], el *TORMES*, el *Loadstone GPS*, el *Personal Guidance System (PGS)* [20], etc.

El dispositivo MoBIC está compuesto por el Pre-Journey System (MoPS), y por el Outdoor System (MoODS), que tiene como objetivo interpretar planos y navegar. Mediante el teclado, el usuario introduce su consulta sobre la ruta o actividad y el MoBIC, mediante sonidos, le da la respuesta [21-22]. El Personal Guidance System (PGS) tiene como objetivo proporcionar navegación a las personas invidentes tanto en entornos conocidos como en

entornos nuevos y desconocidos, sin ayuda de un guía [23]. El rango de trabajo del PGS es de 20 metros.

Entre los dispositivos basados en estereovisión encontramos al Navigation Assistance for Visually Impaired (NAVI), que es un dispositivo detector de obstáculos tanto estáticos como en movimiento lento [24]. Otro dispositivo interesante es el System for Wearable Audio Navigation (SWAN), que también permite la detección de obstáculos [25].

La principal desventaja de la mayoría de Dispositivos Electrónicos de Ayuda a la Movilidad para personas invidentes radica en que no proporcionan una información clara y en tiempo real. La información proporcionada a los usuarios no es suficiente para guiar al usuario de forma independiente y con seguridad, tanto en entornos cerrados como abiertos, comunica Hans Kaltwasser de la Deutscher Blinden-und Sehbehindertenverband e.V. (DBSV).

Partiendo de las limitaciones de los dispositivos desarrollados hasta el presente: distancias cortas, sistema táctil, vibro-táctil (requiere un dispositivo adicional y da incomodidad al usuario), el sistema vocal (requiere mucho tiempo para informar al usuario), o sonidos acústicos (que son sonidos molestos) de los dispositivos existentes todavía a nivel de prototipo así como de las necesidades de las personas invidentes (Seguridad: evitar contacto con los obstáculos, desvío inseguro, no detectar o fallar a la hora de detectar escalones, etc.; Eficiencia: velocidad de caminar, continuidad del viaje, errores de navegación, desvíos de la trayectoria; frecuencia del viaje; distancia del viaje, familiarización y dificultad del viaje; Especificaciones del dispositivo: detección de signos, tiempo de búsqueda, tiempo de lectura, detección de estructuras de superficie y/o texturas, decrecimiento de la precisión de detección, determinación de dirección, determinación de distancia; el ratio de independencia del dispositivo; detección de objetos en entornos reales), los autores se marcaron el objetivo de desarrollar un dispositivo detector de obstáculos y navegación, basado en la tecnología láser, desarrollada para detección de obstáculos, evitar atropellos y accidentes para automóviles. El dispositivo finalmente desarrollado fue denominado Cognitive Aid System for Blind People (CASBlIP). El CASBlIP detecta los obstáculos e indica su localización y distancia precisa tanto de los obstáculos pequeños como de los obstáculos grandes, escalones o pasillos libres [26-27]. La información proporcionada no interfiere con los ruidos del entorno y no se transmiten datos erróneos [28]. Además la información proporcionada es clara y simple, rápidamente interpretada y sin necesidad de entrenamiento excesivo [29-30].

3.2 Arquitectura del dispositivo

El principio de funcionamiento del dispositivo CASBliP se basa en el sistema cognitivo humano; el dispositivo electrónico escanea el entorno y el cerebro humano interpreta la dicha información.

El CASBliP está compuesto por un par de gafas en la que se ha implementado el sistema láser 3D-CMOS (Three Dimensional Complementary Metal Oxide Semiconductor), un FPGA (Field Programmable Gate Array) y un auricular estéreo (Figura 3.1). En la Tabla 3.1 se presentan las especificaciones técnicas del láser 3D-CMOS.

Tabla 3.1 Propiedades técnicas del sensor 3D-CMOS.

Parámetros	Valor
Tiempo de selección	>30ns
Ruido	< 4W/m ²
Área de visión	64º
Sensor utilizado	64x4 3D-CMOS utilizado para el tráfico, el sistema utiliza <i>sólo 64x1lasers</i>
Rango de medida de distancia	De 0,5 m a 5 m
Precisión de medida	< 1% para 100% de información en distancia
Registro del píxel	5 MHz
Energía del pulso de laser	50 µWs
Tiempo de medida	50ms
Tensión de suministro necesaria	12 V
Tecnología del sensor	0,5µm Standard CMOS
Geometría del píxel	130x300µm ²
Longitud de onda del láser	850-910 nm
Interfaz de comunicación On-Board	LAN, RS232



Figura 3.1 Imagen de Cognitive Aid System for Blind People.

3.3 El sistema láser y el método de medida de las distancias

El sistema sensor está compuesto por 1x64 pixeles del láser miniaturizado implementado en unas gafas de sol adaptadas al circuito. La información recibida por los láseres se transmite a una memoria Correlated Double Sampling (CDS), en la que la distancia entre el sistema y los objetos se calcula mediante el método Time-of-Flight (ToF), y dicha información se transmite al sistema acústico. El usuario, mediante auriculares estéreo, percibe la posición en distancia y azimut (dirección) de los obstáculos situados en el área de visión del sistema.

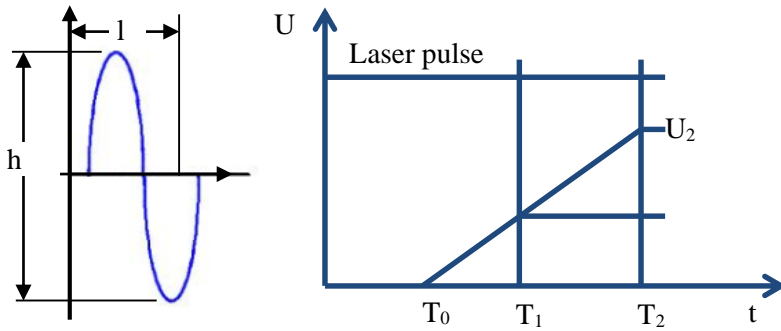


Figura 3.2 Principio de cálculo del área de sensibilidad del píxel y el diagrama de tiempo del método de medida de distancias Time-Of-Flight.

El sistema permite obtener las medidas indirectas de los impulsos Time-Of-Flight del láser en magnitudes del orden de nano segundos (Figura 3.2). Sea h la amplitud y l la longitud de la señal x . Para calcular el tono del píxel y , hay que multiplicar la longitud de la señal por el número de píxeles utilizados n , por el número niveles en elevación m y por la amplitud de la señal h , de donde se obtiene:

$$U_2 \approx E_{laser} * T_p. \quad (3.1)$$

El tono del píxel es $130 \mu m$ en horizontal y $300 \mu m$ en dirección vertical. Debido a que el sistema 3D-CMOS tiene solamente una línea de láseres, (1) se calcula como:

$$y = 64 * 130 \mu m * 1 * 300 \mu m = 2,496 mm^2. \quad (3.2)$$

Cada píxel está compuesto por un foto diodo PD n-well/p-substrate una capacidad C_D , una capacidad de sentido C_{sene0} , una capacidad de reserva C_{HD} , un conmutador de reajuste $\Phi 1$, un conmutador de selección $\Phi 3$, un búfer SF1, un conmutador de conexión $\Phi 4$ y un conmutador de binning $\Phi 2$ (Figura 3.3).

El método de funcionamiento del circuito se basa en que la capacidad del fotodiodo C_D y la capacidad de sentido C_{sene0} periódicamente se reinician al potencial U_{ddpix} obteniéndose una descarga. El tiempo de integración de la descarga está controlado por el conmutador de selección $\Phi 3$. La capacidad C_{HD} lee el potencial restante almacenado en el C_{sene0} . Cuando el conmutador de conexión $\Phi 4$ se conecta, el potencial almacenado en el C_{sene0} se lee mediante la memoria CDS. Al mismo tiempo que el potencial se lee por la capacidad C_{HD} del C_{sene0} , se procede a la lectura del siguiente potencial. De esta forma el circuito realiza una lectura y almacenamiento de los potenciales prácticamente en tiempo real. Para calcular las distancias entre el láser y el objeto, se ha utilizado el principio de medida Time-Of-Flight (TOF) mediante el Multiple Short Time Integrator (MDSI), así como un amplificador de la capacidad de conmutación análogo con la operación Correlated Double Sampling (CDS) [31]. El tiempo de propagación hasta el objeto y de vuelta al sensor 3D- CMOS, se calcula teniendo en cuenta la cantidad de radiación medida en el sensor, la reflexión con la superficie del objeto, la distancia entre el objeto y el emisor y, por fin, la radiación que incluye otras fuentes de luz del entorno. Para poder calcular el tiempo y la distancia es imprescindible utilizar dos tiempos de integración. Sea T_P el tiempo de propagación de la luz de láser, y un T_I el tiempo del integrador corto en el conmutador. Debido a que tenemos dos tiempos sincronizados, el tiempo T_I es igual al tiempo T_P . En la segunda

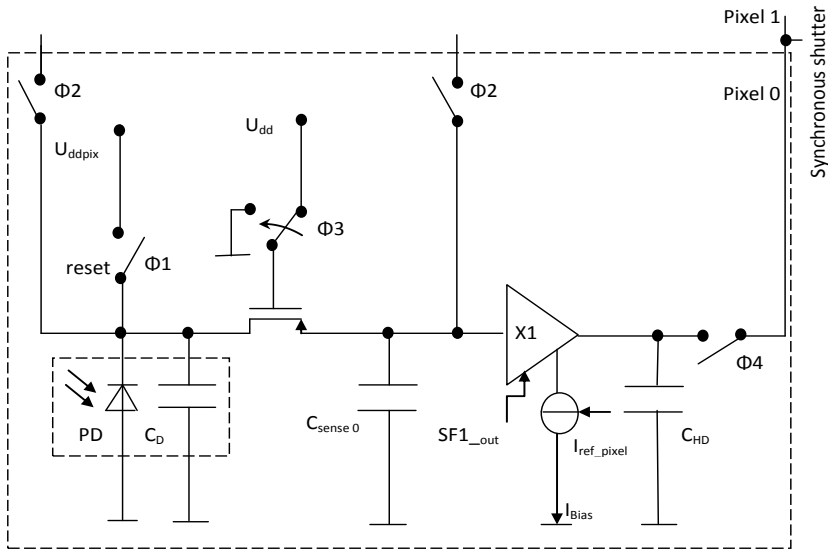


Figura 3.3 Circuito del píxel.

medida tenemos el tiempo T_0 que representa el tiempo medido del impulso de láser reflejado en la superficie del objeto.

Teniendo el tiempo T_0 se puede calcular el tiempo de propagación del impulso de láser:

$$\tau_{TOF} = T_0 = 2\frac{d}{v}. \quad (3.3)$$

Donde d es la distancia medida entre el objeto y el sensor, y el v es la velocidad de la luz.

En el tiempo T_0 se obtiene un crecimiento del potencial U (ver Figura 3.3).

El potencial registrado en el tiempo T_1 es:

$$U_1 \approx E_{laser} * (T_1 - T_2). \quad (3.4)$$

Donde U_1 se guarda en la memoria CDS, y E_{laser} representa la radiación medida en el sensor. Para eliminar los factores externos, así como reflexión de los objetos, iluminación de fondo, etc., se mide un segundo tiempo T_2 denominado tiempo de conmutación de la luz. El tiempo $T_2 \geq 2T_P$

En el tiempo T_2 el potencial registrado es:

$$U_2 \approx E_{laser} * T_P. \quad (3.5)$$

Multiplicando los coeficientes del (4) y (5) se obtiene el valor receptivo y refractivo. Teniendo en consideración que el tiempo $T_1 = T_P$ se obtiene:

$$\frac{U_1}{U_2} = \frac{(T_P - T_0)}{T_P}. \quad (3.6)$$

De donde se puede calcular el tiempo T_0 :

$$T_0 = T_p * \left(1 - \frac{U_1}{U_2}\right). \quad (3.7)$$

Sustituyendo (7) en (3), se puede calcular la distancia entre cada píxel de los 64 utilizados entre el sistema 3D-CMOS y la superficie del objeto:

$$d = \frac{v}{2} * T_p * \left(1 - \frac{U_1}{U_2}\right). \quad (3.8)$$

3.4 Sistema acústico y método de producción de distancias mediante sonidos

El sistema acústico del dispositivo detector de obstáculos y navegación tiene como objetivo representar mediante sonidos acústicos los 64 valores de distancias enviados por el sistema del sensor 3D-CMOS. Con este fin, es imprescindible hacer un estudio exhaustivo de las propiedades del sonido a utilizar, medir y generar las Funciones de Transferencia, analizar las propiedades psicoacústicas de los sonidos generados y, finalmente, implementar dichos sonidos en el sistema final y verificar su eficacia. Para su funcionamiento, el sistema acústico está compuesto por un banco de sonidos de 64 sonidos ocupando 62° en azimut y 16 niveles en distancia desde 0,5m hasta 5m (ver Tabla 3.2), que aumentan de forma exponencial. Para cada uno de los láseres se le ha atribuido un sonido específico.

Para generar los sonidos a cada distancia, se han utilizado las medidas de las Funciones de Transferencia (HRTF) en una sala anecoica con la ayuda de un robot KEMAR. Un Maximum Length Binary Sequence (MLBS) se utilizó como fuente sonora para la medida de los HRTFs.

Sea $x_I(t)$ la señal reproducida por el altavoz e $y_I(t)$ la señal grabada por el micrófono colocado en el canal audio del maniquí robot. Se diseña un filtro digital para transformar $x_I(t)$ en la señal $x_2(t)$ reproducida por el auricular. Sea $X1$ la representación del $x_I(t)$ en el dominio de la frecuencia, e $Y2$ la

representación de $x_2(t)$. Al reproducir $x_1(t)$, la respuesta que se detecta se puede representar como:

$$Y_1 = X_1HM. \quad (3.9)$$

Donde L representa la Función de Transferencia del altavoz con toda la cadena de audio de reproducción, F es la Función de Transferencia del espacio libre entre el altavoz y el canal auditivo, y M es la Función de Transferencia del micrófono y toda la cadena de audio de grabación. De la misma forma se calcula la respuesta grabada por el micrófono al reproductor $x_2(t)$.

$$Y_2 = X_2HM. \quad (3.10)$$

Donde H representa la Función de Transferencia del auricular al canal auditivo. Si $Y_1=Y_2$ obtenemos el filtro digital:

$$T = \frac{LF}{H}. \quad (3.11)$$

Finalmente, al someter el sistema a una Secuencia Binar de Máxima Longitud (MLBS) que denominamos $s(n)$, la respuesta del impulso periódica $h(n)$ será:

$$h(n) = \Omega_{sy}(n) = s(n) * \Phi y(n). \quad (3.12)$$

Donde Φ es la correlación cruzada circular periódica. Para reducir el tiempo de cálculo se utiliza la Transformada Hadamard Rápida (FHT).

$$h(n) = \frac{1}{(L+1)s[0]} P_2 \langle S_2 \{ H_{L+1} [S_1 (P_1 y(n))] \} \rangle. \tag{3.13}$$

Donde P son las matrices de permutación y las S-las matrices de redimensionamiento, HL+1 representa la matriz Hadamard de orden L+1. Para realizar la convolución con el sonido se ha utilizado la Transformada Fourier Discreta. En la Tabla 3.3 se presentan las características técnicas del sonido final.

Durante las pruebas experimentales participaron 20 personas (19 personas ciegas y una persona con visión reducida (casi nula) con edades comprendidas entre 26 y 69 años, tanto hombres como mujeres. Diez de los individuos eran de Alemania y diez de Italia. Todos poseían una buena capacidad auditiva y eran capaces de estimar la distancia y la dirección de los objetos. El grupo de usuarios abarcaba desde ciegos de nacimiento hasta personas con pérdida de visión entre 0 y 5 años debido a enfermedades o accidentes. El grupo de usuarios ha sido seleccionado por las asociaciones respectivas de Italia y Alemania, así como supervisados por los entrenadores de las dichas asociaciones.

Para las pruebas de aprendizaje se realizaron durante una semana siete ejercicios con distinto nivel de complejidad, desde la percepción de un simple obstáculo hasta la percepción de un obstáculo frente a una pared u otro obstáculo.

Tabla 3.2 Número de profundidad y las distancias correspondientes.

Nº of Depth	Distance to the user in cm.	Nº of Depth	Distance to the user in cm.
1	0,500	9	1,707
2	0,583	10	1,991
3	0,680	11	2,321
4	0,792	12	2,706
5	0,924	13	3,155
6	1,077	14	3,678
7	1,256	15	4,288
8	1,464	16	5,000

Después de superar con éxito las pruebas de aprendizaje, los participantes han pasado a la segunda prueba de navegación, denominada prueba de laboratorio. La prueba de laboratorio consiste en un ejercicio de detección de obstáculos situados en forma de un pasillo de 14 m, haciendo uso de auricular. Utilizando cajas de cartón se han creado 4 pares de obstáculos colocados en paralelo a una distancia de 2,5m.

Durante la prueba, los participantes debían mover la cabeza a derecha y a izquierda para detectar los obstáculos y encontrar el pasillo libre.

En las pruebas en entorno real se han utilizado una variedad de emplazamientos: un patio del instituto (un lugar bien conocido por los participantes, pero con obstáculos construidos a base de cajas de cartón), una calle y, finalmente, un centro comercial.

Tabla 3.3 Características técnicas del banco de sonidos.

Parámetros	Valor
Dimensiones Azimut del campo de visión del dispositivo de audio	62°
Distancia del campo	Entre 0,5 hasta 5 m
Intervalo mínimo de tiempo entre sonidos	8 ms
Número de píxeles reproducidos por imagen	64
Frecuencia de refresco	2 frames/sec
Longitud máxima de los sonidos	2048 muestras
Frecuencia de muestra	44.100 Hz
Tamaño de la colección de sonidos	16 Mb
Generación de las secuencias de sonido	Aleatoria
Control del nivel de sonido del sistema audio	8 niveles

3.5 Resultados

La validación del dispositivo electrónico detector de obstáculos mediante el láser 3D-CMOS y sonidos acústicos se realizó mediante unas pruebas de aprendizaje, una prueba de laboratorio y mediante unas pruebas en entorno real. Las pruebas de aprendizaje tenían como objetivo el aprendizaje de percepción de sonidos acústicos mediante auriculares, esto es, la percepción de que los dichos sonidos provienen desde el entorno real y no de los auriculares, percibir las formas de los objetos que se crean mediante los sonidos acústicos y percibir la dirección y la distancia entre los sonidos y el usuario. Las pruebas se realizaron con el dispositivo electrónico, tanto con ayuda como sin ayuda del bastón.

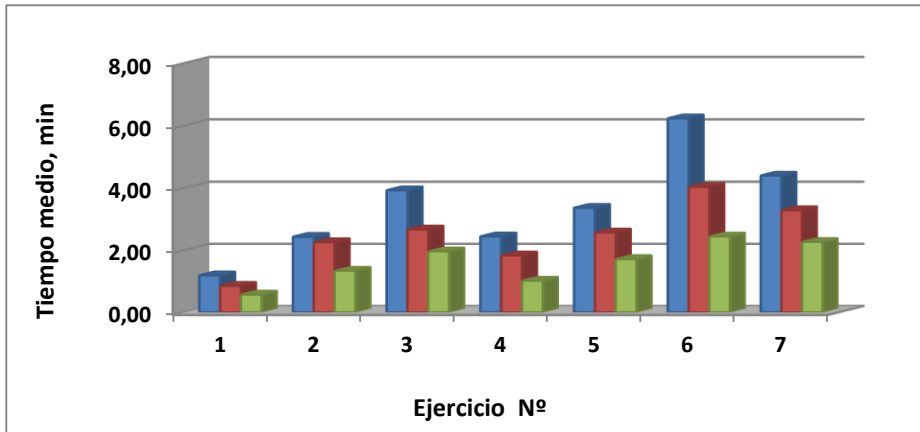


Figura 3.4 Diagrama de tiempo de los siete ejercicios de la prueba de aprendizaje. El eje x representa cada uno de los siete ejercicios; los valores del eje y representan el tiempo en minutos; las barras en colores representan las tres pruebas realizadas para cada ejercicio.

En la Figura 3.4 se representan los resultados obtenidos de los siete ejercicios realizados por las personas invidentes de DBSV. Se puede observar importantes diferencias entre las tres pruebas. El promedio de tiempo de la primera prueba es de 3,29min, cuando la segunda prueba el resultado mejor

con 1,31min, y la diferencia entre el segundo con el tercer resultado es de 1,65 min.

Los resultados de las pruebas, reflejados en la Figura 3.5, revelan que el tiempo necesario para realizar el recorrido aumentó considerablemente cuando la prueba se hizo sin ayuda del bastón.

Comparar los resultados obtenidos con ayuda del bastón y los resultados conseguidos únicamente con ayuda del dispositivo electrónico no es uno de los objetivos del trabajo, ya que todos los participantes en el experimento tenían mucha experiencia en el uso del bastón y casi nula con el dispositivo.

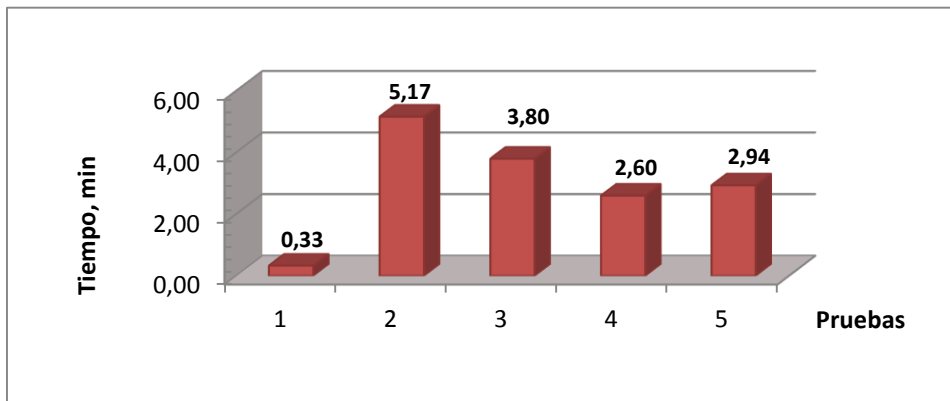


Figura 3.5 El diagrama de tiempo de las 5 pruebas de laboratorio. El eje x representa cada una de cinco pruebas (la prueba 1 representa el resultado con ayuda del bastón, las siguientes dos columnas representan los resultados promedio de ida y vuelta del ejercicio de los diez participantes de Alemania, mientras que las columnas 4 y 5 representan el promedio de los resultados de ida y vuelta del grupo de Italia). Los valores del eje y representa el tiempo en minutos.

Sin embargo, sí que se observaron mejoras respecto a la repetición de la prueba, lo que constituye un aliciente para seguir con la idea de que el dispositivo resulta viable y puede ser utilizado como método de navegación. Otra observación interesante es que, dado que los obstáculos estaban situados a una distancia cercana, la detección mediante bastón resultaba más fácil si cabe.

En todos los casos los participantes han demostrado una gran habilidad de navegación con el dispositivo teniendo en cuenta el tiempo reducido de práctica (sólo un mes). Es de destacar que, comparando todas las pruebas, la velocidad de navegación obtenida con el dispositivo acústico es la mitad que la velocidad de navegación utilizando el bastón. Esto se debe a la práctica de los participantes con el bastón. Sin embargo, se ha demostrado que utilizando el dispositivo electrónico desarrollado, los participantes ciegos eran capaces de percibir y detectar los obstáculos lejanos y no solamente los ubicados a un metro de distancia (lo cual es posible mediante bastón). Además el dispositivo les facilitaba la percepción de lo que les rodeaba, distinguiendo si se trataba de obstáculos estáticos o en movimiento. También les facilitaba el movimiento, ya que al saber lo que les rodeaba en un área de cinco metros cuadrados, los usuarios eran capaces y libres de tomar decisiones de seguir adelante o cambiar de trayectoria. En combinación con el bastón en entornos conocidos, los participantes podrían fácilmente identificar los obstáculos (por ejemplo, “ésta debe ser la farola” o “éste debe ser el kiosco”) según se perciba el volumen asociado a los mismos. El entorno en el que la orientación ha resultado más complicada ha sido en un centro comercial lleno de gente, donde al principio los participantes se quedaban desorientados por la cantidad de información percibida.

La ventaja del dispositivo es la cantidad de información recibida mediante los auriculares que se transmite en tiempo real con una precisión de $0,97^\circ$ en dirección y una distancia de 0,30 cm en distancia, gracias al sensor 3D-CMOS de gran precisión y a los sonidos acústicos empleados de gran calidad psicoacústica para reducir el tiempo de reproducción e incrementar la precisión y la suavidad del sonido.

3.6 Conclusiones

En el presente artículo se ha presentado un nuevo dispositivo de navegación para las personas invidentes, basado en la tecnología láser y acústica. El dispositivo acústico desarrollado tiene la capacidad de detectar todos los obstáculos con una precisión de 99% para el sistema 3D-CMOS y un 95% para el sistema acústico. Al combinar las dos tecnologías, los usuarios ciegos son capaces de percibir y localizar todos los obstáculos que aparecen en el área de visión del dispositivo, desde 0,5 metros hasta 5 metros en distancia y 60° en dirección. El dispositivo desarrollado es capaz de transmitir en tiempo

real al usuario toda la información sobre los obstáculos del entorno, de forma que el usuario tiene la posibilidad de tomar decisiones sobre cómo actuar. Con el Dispositivo Acústico desarrollado los resultados obtenidos han sido realmente impresionantes, los usuarios han sido capaces de detectar los obstáculos y esquivarlos en un par de horas de entrenamiento. Han sido capaces de percibir los sonidos del dispositivo como si provinieran desde el entorno y no desde los auriculares. Han sido capaces de percibir las formas y el material de los obstáculos (plano o con deformaciones, de madera, cemento o metal...).

3.7 Agradecimientos

El trabajo ha sido apoyado por el proyecto N° 2062 del programa PAID-06-11 de la Universitat Politècnica de València.

3.8 Referencias

- [1] WBU (World Blind Union) Available:
<http://www.who.int/mediacentre/factsheets/fs282/en/> [Citado 28 de Abril 2013]
- [2] DBSV (Asociación alemana de ciegos y discapacitados visuales).
<http://www.dbsv.org/infothek/zahlen-und-fakten/> [Citado 20 de Junio 2012].
- [3] ChinaDis Available:
<http://disabilitychina.org/admin/upLoadPic/2009326152139974.pdf>
[Citado 10 de Mayo 2012].
- [4] Starkiewicz, W. and Kuliszewsky, T. Active energy radiating system: the 80-channel Electroftalm. Proceedings Int. Congress on Technology and Blindness. Vol. 1. 157-166, 1963.
- [5] Nye, P.W. and Bliss, J.C., Sensory aids for the blind: A challenging problem with lessons for the future. Proceedings of the IEEE. Vol. 58-12. pp. 1878-1898, 1970.
- [6] Capp, M., and Picton, P.D., The optophone: An electronic blind aid. Engineering Science and Education Journal. Vol. 9-3. pp. 137-143, 2000.

- [7] Dakopoulos, D. and Bourbakis, N.G., Wearable Obstacle Avoidance Electronic Travel Aids for Blind: A Survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 40, 1, pp. 25-35, 2010.
- [8] Dunai, L., Peris Fajarnes, G., Santiago Praderas, V., Defez Garcia, B., Electronic travel aid systems for visually impaired people. *Proceedings of DRT4ALL 2011 Conference IV Congreso Internacional de Diseño, Redes de Investigación y Tecnología para Todos*, Madrid, Spain. 2011.
- [9] Quintià Pardo, H., Calvo Role, J. L., Fontenla Romero, O. Aplicación de un robot comercial de bajo coste en tarea de seguimiento de objetos. *Dyna*, 79(175), pp.24-33, 2012.
- [10] Russell, L. Travel path sounder. *Proceedings. Rotterdam Mobility Research Conference*. N.Y: American Foundation for the Blind. 1965.
- [11] Mann, R.W. Mobility aids for the blind—An argument for a computer-based, man-device-environment, interactive, simulation system. *Proceedings of Conference on Evaluation of Mobility Aids for the Blind*, Washington, DC: Com. on Interplay of Eng. with Biology and Medicine, National Academy of Engineering. 101–116. 1970.
- [12] Morrisette, D.C., Goddrich, G.L., Henesey, J.J., A follow-up-study of the Mowat sensor's applications, frequency of use, and maintenance reliability. *J. Vis. Impairment Blidness*. Vol. 75-6, 244-247, 1981.
- [13] Loomis, J.M., Golledge, R.G., Klatzky, R.L., GPS-based navigation systems for the blind. In W. Barfield & T. Caudell Eds. *Fundamentals of wearable computers and augmented reality*. Mahwah, NJ: Erlbaum. pp. 429-446,2011.
- [14] Kuc, R., Binaural sonar electronic travel aid provides vibrotactile cues for landmark, reflector motion and surface texture classification. *IEEE Transactions on Biomedical Engineering*. Vol. 49-10, pp 1173-1180, 2002.
- [15] Farmer, L. and Smith, D., Adaptive Technology. Book chapter, *Foundations on Orientation and Mobility*. Eds. 2. pp. 231-260, 1997.
- [16] Farmer, L.W., Mobility devices. *Bulletin of Prosthetic Research*. 47-118, 1978.

- [17] ANDO, B., Electronic sensory systems for the visually impaired. *Instrumentation and Measurement Magazine, IEEE*, Vol 6, (2), pp. 62-67, 2003.
- [18] Petrie, H., Johnson, V., Strothotte, T., Raab, A., Michel, R., Reichert, L., Schalt, A., MoBIC: An aid to increase the independent mobility of blind travellers. *British Journal of Visual Impairment*. Vol. 15-2, pp. 63-66, 1997.
- [19] Loomis, J., Re: Digital Map and Navigatio System for the Visually Impaired. *Comunicación personal*, 1985.
- [20] Loomis, J.M., Golledge, R.G., Klatzky, R.L., GPS-based navigation systems for the blind. In W. Barfield & T. Caudell Edss. *Fundamentals of wearable computers and augmented reality*. Mahwah, NJ: Erlbaum. pp. 429-446, 2001.
- [21] 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.
- [22] Strothotte, T., Petrie, H., Johnson, V., Reichert, L., MoBIC: user needs and preliminary design for a mobility aid for blind and elderly travellers. *British Journal of Visual Impairment*. Vol. 15-2, pp. 63-66,1997.
- [23] Loomis, J.M., Golledge, R.G., Klatzky, R.L., Navigation Systems for the blind:Auditory display modes and guidance. *J. Presence*. Vol 7, No.2, pp. 193-203, 1998.
- [24] Loomis, J.M., Golledge, R.G., Klatzky, R.L., AND Marston, J.R., Assisting wayfinding in visually impaired travellers. In G. Allen (Ed.). *Applied spatial cognition: From research to cognitive technology*. 179-202. Mahwah, N. J.: Lawrence Erlbaum Associates, 2007.
- [25] Sainarayanan, G., Nagarajan, R., and Yaacob, S., Fuzzy image processing scheme for autonomous navigation of human blind. *Appl. Soft Comput. Journal*. Jan. Vol. 7-1 pp. 257-264, 2007.
- [26] Wilson, J., Walker, B.N., Lindsay, J., Cambias, C., Dellaert, F., SWAN: System for wearable audio navigation. *11th IEEE International Symposium on Wearable Computers (ISWC'07)* Boston. MA. pp. 91-98, 2007.
- [27] Benham, T.A., Guidance Device for the Blind. *Phys. Today*. Vol. 7-12, pp. 11-14, 1954.

- [28] BENJAMIN, J.M. A review of the veterans administration blind guidance project. Bull. Of Prosth. Res. 63-90,1968.
- [29] Dupress, J.K., The requirements for successful travel by the blind. In L. L. Clark Eds. Vol. 1. pp. 7-11,1963.
- [30] Farmer, L.W., Travel in adverse weather using electronic mobility guidance devices. The new outlook for the blind. Vol. 69-10. pp. 433-439, 1975.
- [31] Elkhali, O., Schrey, O.M., Mengel, P., Petermann, M., Brockherde, W., A 4x4 pixel CMOS image sensor for 3-D measurement applications. in IEEE Journal of solid state circuits. Vol. 39-7. pp. 1208-1212, July. 2004.

4. ARTICULO II

Larisa Dunai, Guillermo Peris-Fajarnes, Ismael Lengua, Ignacio Tortajada M.
MATERIAL SOUND SOURCE LOCALIZATION THROUGH HEADPHONES.

Acoustical Physics, 2012, Vol. 58, No. 5, pp. 610–617. ISSN 1063-7710.

Abstract

In the present paper a study of sound localization is carried out, considering two different sounds emitted from different hit materials (wood and bongo) as well as a Delta sound. The motivation of this research is to study how humans localize sounds coming from different materials, with the purpose of a future implementation of the acoustic sounds with better localization features in navigation aid systems or training audio-games suited for blind people. Wood and bongo sounds are recorded after hitting two objects made of these materials. Afterwards, they are analysed and processed. On the other hand, the Delta sound (click) is generated by using the Adobe Audition software, considering a frequency of 44.1 kHz. All sounds are analysed and convolved with previously measured non-individual Head-Related Transfer Functions both for an anechoic environment and for an environment with reverberation. The First Choice method is used in this experiment. Subjects are asked to localize the source position of the sound listened through the headphones, by using a graphic user interface. The analyses of the recorded data reveal that no significant differences are obtained either when considering the nature of the sounds (wood, bongo, Delta) or their environmental context (with or without reverberation). The localization accuracies for the anechoic sounds are: wood 90.19%, bongo 92.96% and Delta sound 89.59%, whereas for the sounds with reverberation the results are: wood 90.59%, bongo 92.63% and Delta sound 90.91%. According to these data, we can conclude that even when considering the reverberation effect, the localization accuracy does not significantly increase.

4.1 Introduction

One of the most important factors for human survival is the ability of sound source localization in real environments. This ability helps humans to avoid obstacles that appear in their way, to prevent dangers, to perceive the differences between different types of sounds as well as to judge their representation, to perceive speech, etc. It is important to remark that the sound source localization in the environment is a difficult task, because of the influence of different acoustical cues, such as reverberation and echo, as well as the huge amount of different noises that are present in this atmosphere.

In this regard, it is important to note that a wide research has been developed on sound source localization [1]. The influence of binaural cues, such as Interaural Time Differences and Interaural Level Differences, on sound source localization has drawn special attention [2-3].

Also the auditory saltation or the audible threshold has been deeply studied, since they are important factors of sound localization [4-7]. Due to the improvement and great demand of acoustical systems and virtual technologies, sound source localization has been also analysed by using generated or measured Head- Related Transfer Functions (HRTF), by means of the creation of virtual environments [8-9].

With regards to generated HRTFs, the main task is the computer model used for the creation of the environment. The main problems of the computer algorithm for architectural acoustic rely on the acoustic characteristics of the room, as well as on how to estimate reverberation time or sound quality [10-11].

It is a great dilemma how human auditory system processes and perceives material sounds. [12-13], investigated the material perception and the influence of variables that govern the synthesis of material sounds. Actually, it is easy to extract acoustic information from the sounds, but the question is how the human brain captures this information and allows us to classify and localize material sound sources. It is important to remark the influence of material physical properties on material sound perception. Among them, the material elasticity, the volume of the object, the force with which the object has been impacted or the frequency are especially important. Up to date, most of the works have intended to explain how humans perceive two different material sounds [12, 14]. However, no works have been found on localization of azimuth and distance of material sounds.

Material sound perception and localization provide very useful information both for representing the surrounding environment and for virtual reality applications. For example, when an impaired person walks on the street, he combines his hearing abilities and his white cane, in order to detect the obstacles in his path and to avoid them. Also, it is crucial for them to perceive the differences between material sounds. This property helps them to analyse the danger of the obstacles and the environment. The sounds or noises generated by cars, street lights, people walking, opening and closing doors, etc., are important survival cues for visually impaired people. On the other hand, with the development of new acoustic navigation aids for blind people, the selection of the correct acoustic sounds becomes a primordial task. This is due to the fact that the impaired user uses the device during long time periods and, therefore, the implemented sounds must be clear, short, and not interfere with the surrounding noises in order to avoid irritating him. In conclusion, the user safety strongly depends on the selected sounds.

With all these facts in mind, the current study aims to analyse the sound source localization, considering three different sounds: wood, bongo and a Delta sound generated using HRTFs. The study considers two different situations: an anechoic environment (i.e. no reverberation) and an environment with reverberation. The paper explains the simple psychoacoustic experiment developed, which implies the use of a graphic user interface as well as of previously measured HRTF. The reason for selecting wood, bongo and Delta (or click) sounds in this psychoacoustic experiment is motivated by their widely spread presence in the human everyday life; these sounds are very common to the listeners, being very their discrimination. Also, it is important for the psychoacoustic experiments to know the effects of the environmental conditions, such as the reverberation (the importance of which has been pointed out by many authors [15]), which enhances the realism of the sounds, providing a more realistic sensation. The paper analyses the responses obtained in all the categories.

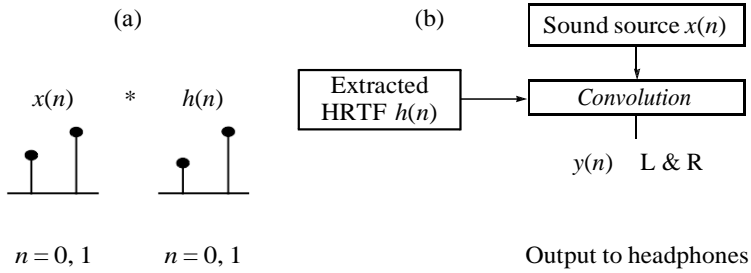


Figure 4.1 A convolution representation of the two sequences. In (a), it is shown the mathematical representation of the convolution, in (b) the convolution algorithm between the HRTF and the previously generated signal is presented.

Table 4.1 Time duration of the sounds.

Sound	Time in seconds	
	Anechoic	Reverberation
Wood	0.078	0.548
Bongo	0.104	0.548
Delta (click)	0.078	0.548

4.1.1 Procedure

Sound Sample

Two different sounds are firstly obtained by hitting everyday life objects made of wood and bongo. Besides, a Delta sound (a click) is generated with the Adobe Audition software. Then, each sound is simplified based on additive synthesis techniques that resample these recorded sounds at frequency of 44.1 kHz and sound level of 72 dB. All sounds are recorded and generated in two different environments: anechoic and with reverberation. Each signal has different spectral content and duration (See Table 1). A reverberation effect is applied to each sound by using the CATT-Acoustic software with FireReverb™ Suite by PureVerb™. The environment is

created considering the following room dimensions: volume $V = 16 \text{ m}^3$ (width \times length \times height = $2 \times 4 \times 2 \text{ m}^3$), and the following values for the program parameters: absorbing element $A = 0.25 \text{ s}$, reverberation time $T = 1.4 \text{ s}$ and directivity index (which controls the direct sound to reverberation ratio) $D = 1.0 \text{ dB}$. Finally, each sound is convolved with a non-individual Head-Related Transfer Function (See Figure 4.1). The HRTF is generated using a KEMAR dummy head. It is measured for each spatial position in a frontal plane, both in distance and in azimuth:

$$y(n) = x(n)*h(n), \tag{4.1}$$

where $x(n)$ represents a sound with n samples and $h(n)$ represents the Head-Related Transfer Function of n samples.

In the convolution process between the HRTFs and the sound, it can happen that the convolution is obtained only for one sound channel. Applying this method, the sounds are converted into monaural

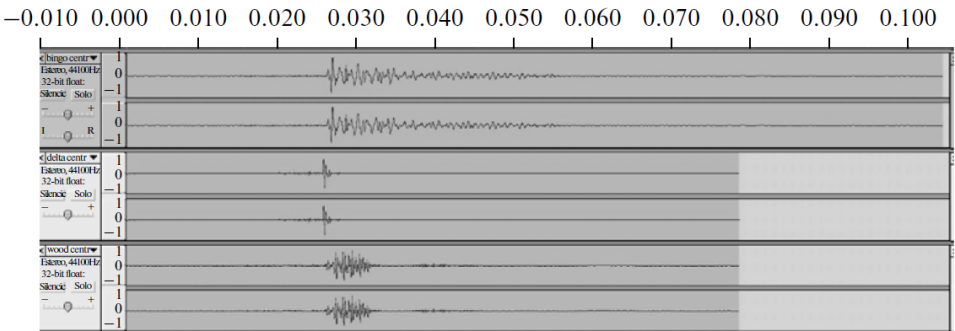


Figure 4.2 Representation of sound waves obtained after the convolution between the bongo, Delta sound (click) and wood sounds with the HRTF, for a distance of 0.5 meters and direction of 0° , using the Discrete Fourier Transform method. The x -axis represents the time in seconds (the bongo sound lasts 0.104 s, the delta and wood sounds last 0.078 s), while the y -axis represents the wave amplitude after convolution and normalization (between -1 and 1 values). Each of the three sounds is composed by the two channels (left and right) of the stereo sound.

sounds. Since the objective of this experiment is to create binaural sounds, (1) cannot be used alone. The solution is to apply the Discrete Fourier Transform (DFT) (2), which allows us to transform the signal from time domain to the frequency domain.

$$x(k) = \sum_{n=0}^{N-1} x(n)e^{\frac{-jnk2\pi}{N}} . \quad (4.2)$$

Where $x(k)$ represents the signal resulting from the DFT and $x(n)$ is the signal convolved. By using DFT (2) we obtain a sequence of complex numbers with length $N/2+1$; the output signal contains two channels. Both resulting signals will contain the signal magnitude. Applying the convolution with the HRTFs, the sound is spatialized to appear as if it was located at the corresponding real-world location. This method is used to obtain 3D-stereo sounds to be employed in experiments using headphones. In this way, the sounds listened through the headphones appear as if they were inside the head. In order to represent sounds as if they were coming from the environment, it is important to convolve them with previously generated HRTFs, by using a proper method. In this paper, the convolution has been performed, by using Discrete Fourier Transform.

In order to generate the HRTFs, the Maximum Length Binary Sequences MLBS is applied and later a cross correlation method is used between the system answer and MLBS input. Each sound is processed using Matlab and saved as a Microsoft “.wav” file sampled at 44.1 kHz and with a 16-bit resolution (Figure 4.2).

4.2 Listeners

Ten young volunteers take part in the experiment. The average age is 25 years. All of them report having normal hearing. Listeners are previously tested with Békésy technique [Gelfand S.A., 2009], [Jerger J., 1962] using pure tones between 250 and 8000 Hz. The threshold is within 15 dB in both ears. The technique consists of recording the auditory threshold on an auditory

blank. The technique is based on two motors: one motor increases the sound frequency slowly, while the second one causes the tone to become gradually louder. When the listener hears the sound, he pushes the button. In this case, the sound becomes gradually fainter. When he does not hear any sound, he releases the button and the sound gets gradually louder. During the test the frequency is increasing slowly from a very low frequency. All of them are experienced in sound localization tasks.

4.3 Method

Stimuli are presented to the users through SENNHEISER HD201 headphones, which are connected to an ACER laptop. For this purpose, an experimental program developed in Flash 8.0 and running on Windows XP is used. The program enables the reproduction of qualitative sounds. The subjects interacted with the experiment using the graphic user interface depicted in Figure 4.3.

This experiment is conducted in a conventional laboratory, in which external noises are presented (human speech, noises coming from the street, equipment noises, etc). The experimental interface consists of a representation of an horizontal plane, corresponding to distances between 0.3 m and 5 m and azimuths between 75° (left) and 75° (right). It is divided into 17 areas in distance and in 27 areas in azimuth, resembling a grid. In the top of the interface, sound location data, such as distance, azimuth and mean are specified. At the left side of the interface, the subject can find the 'Repeat the Sound' button. This button allows the subject to repeat the previously listened sound in case he was unable to hear it due to any reason. Below the 'Repeat the Sound' button, a counter of remaining sounds is placed.

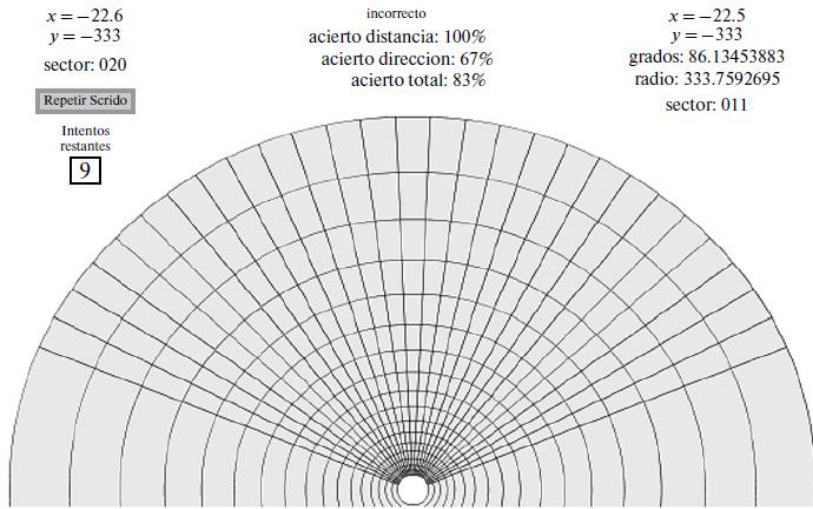


Figure 4.3 Graphic user interface, where the upper side of the interface represents the coordinates of the mouse, the results are in percent and the coordinates of the selected sound and the number of rested sounds that are to be reproduced.

The experiment is distributed in two groups: the anechoic material sounds and material sounds with reverberation. Each group consists of three sounds: wood, bongo, and delta sound. In each trial, the listener hears a train of sounds in a spatial position randomly delivered by the program and estimates its position, without trying to guess the material. After the user has answered, the next train of sounds is delivered. Each trial consists of ten sounds reproduced in different positions. The method uses the forced choice approach, this is, a sound is presented to the user in a certain location and the listener is required to report whether the sound has appeared at the left, right, center, near or far by pressing one cell of the user interface. Even if he did not understand or listen the sound, he is forced to answer. In other words, ‘I do not know’ is not an acceptable response. In the experiment, the duration of the sound is not taken into account. Participants can repeat the sound as many times as they decide by pressing the REPEAT button. Then, the listener responds by clicking with the mouse on one of the cells on the interface. The volume control is constant for all subjects. There is no feedback. A typical run lasts 5-8 min, depending on the listener.

All listeners perform one trial for each of the six types of sounds. Firstly, the listeners performed the experiments for each of the three types of sounds in the anechoic environment and afterwards in the reverberation environment. Provided that the program delivers only one material sound type per trial, it is assumed that this method does not affect participants' response. The program delivers the sounds, one by one, in a randomized order for each stimuli and trial. After hearing the sound through the headphones, the listeners are asked to click on the cell from which they consider the sound has come. Before giving the response, the subjects are allowed to repeat the stimulus. On each trial, the responses are given and recorded by clicking the mouse button pointing the cursor in one position of the cell interface.

4.4 Results

The psychoacoustic experiment validates the localization of the source of material sounds via headphones. By means of the Head Related Transfer Functions, the sound obtains the effect of 'free field'; this means that the sound appears as if it comes from the surrounding instead of from the headphones. The acoustical bases of the localization of material sounds processed both in the anechoic environment and in the environment with reverberation are investigated.

Sound source position perception is defined by azimuth angle and distance. The azimuth errors are defined as the angular difference between the defined sound source and the perceived sound source. Distance errors are defined in an analogous way.

The first step when analysing the data is to determine the mean response for each listener in each run. Afterwards, all the mean responses are collected and the global mean for all listeners calculated. Figure 4.4 illustrates the mean and standard deviation corresponding to material sound localization both for anechoic and for reverberate sounds. The standard deviation (N-1) is calculated from the mean responses for each sound in both conditions.

$$sd = \sqrt{\frac{\sum (x - \bar{x})^2}{(n - 1)}}. \quad (4.3)$$

Where x is the localization mean \bar{x} represents its average and n represents the total runs number.

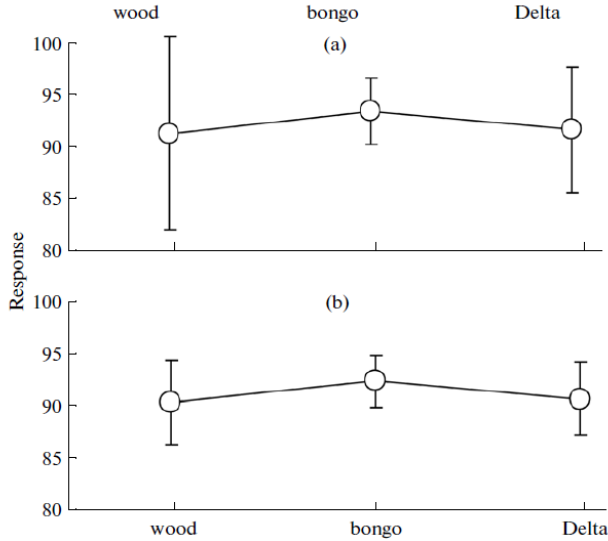


Figure 4.4 Average localization responses for ten listeners plotted as a function of material wood, bongo and Delta (click) in %, where (a) represents the results for the anechoic sounds and (b) represents the results of the sounds with reverberation. Errors bars are standard deviation.

Figure 4.4 Shows that there are no great differences between anechoic and with reverberation localization accuracy but great differences appear in the standard deviation. Note that the standard deviations are smaller for the responses corresponding to sounds with reverberation.

Table 4.2 and Figure 4.4 prove that mean values for wood and Delta sounds with reverberation show slight improvements in comparison with the anechoic sounds. This demonstrates that the reverberation effect improves the subjective realism. Again, we can confirm that reverberation effect helps humans on sound localization, at least in distance [Mershon et al., 1989].

Table 4.2 Mean and minimum values in percentage for all three rounds for anechoic and reverberant situations.

Value, %	Anechoic			Reverberation		
	wood	bongo	Delta	wood	bongo	Delta
Maximum	97.1	96.6	95	96.7	95.4	94.9
Mean	89.87	93.38	88.61	91.63	92.83	91.11
Minimum	71.5	88.4	80.5	83	87.7	83.1
Mean <i>sd</i>	8.8	3.1	5.8	4	2.5	3.5

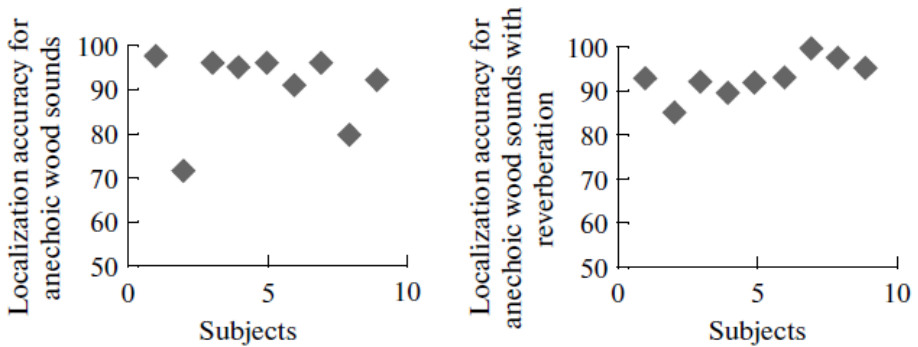


Figure 4.5 Localization accuracy for wood sound (anechoic and with reverberation). The x-axis represents the averaged localization data by subjects and y-axis represents the localization accuracy in %.

On the other hand, it is difficult to reach the same conclusion for the “bongo” sound, in which data differ only in one percent between anechoic and with reverberation sounds.

Also, small differences can be observed in the standard deviations when comparing the anechoic and with reverberation bongo and delta sounds, but for wood sounds the standard deviation dramatically increases. It is observed that the deviation for the anechoic wood sound is more than twice higher than for the environment with reverberation. This means that the reverberation effect provides a higher precision in the perception judgments. On the other

hand, the anechoic sounds are, due to their inherent physical properties, poor, leading to troubles in the localization task, especially for the wood material sound. Fig. 5 shows the mean results and the differences for the ten listeners which take part in the experiment for wood sound localization. From Fig. 5 we can observe that listener 2 is one of the not successful listeners, he has troubles in the localization task in both the anechoic and with reverberation sounds. For this user, the wood sound was confusing and less clear. Another interesting conclusion is that the localization accuracy in distance- or depth- is worse than in azimuth for almost all listeners. This fact can be justified because the Interaural Level Difference and Binaural Time Difference give less information regarding distance localization [Kim D.C. et al, 2008]. Also, many non-acoustic properties contribute to the sound localization in distance [Zahorik, 2002].

In most cases, the listener distraction has an important influence on localization accuracy; after several sounds, some listeners become distracted by the environment surrounding the listener or the external noises such as street noises.

One of the listeners, listener X, develops the experiment to analyze the influence of a long training period on the localization task for anechoic material sounds. The listener X is supposed to carry out this test due to his great abilities for sound localization in an anechoic chamber and with headphones as well as to the low Interaural Time Difference threshold obtained by him in previous experiments. Analyzing localization in distance and azimuth, the listener X obtained better results for azimuth material sound localization. Performances on localization and standard deviation (N-1) are presented in Table 4.3.

Taking into consideration the standard deviation (N-1) for listener X in Table 4.3, we can observe that the localization performances in azimuth are more constant than the localization results in distance. This means that the listener perceives better the direction of the sound source than the distance between the sound source and himself. For example for the Delta sound, the standard deviations (N-1) for azimuth and distance localization are very small and similar. In this case, listener X is able to accurately perceive the Delta sound for azimuth and distance with a standard deviation of 2.11%. With the bongo sound source, there is a great variation in the results regarding distance localization; the standard deviation is 11.3%. For the bongo sound source, listener X has troubles with distance localization for the first four runs. We

must mention that, for the bongo sound, an improvement in distance localization is perceived in the first six runs.

In Figure 4.6, we can observe that, during the first four trials, listener X experiences significant oscillations in perception accuracy for distance between material sounds. From the fifth trial, listener X improves his localization accuracy both in distance and in azimuth.

Listeners are able to represent the sound source position with great accuracy.

The present research has direct implications for the design of environments for Virtual Reality in which the interaction with the objects is presented in a similar way to the human visual interaction in the real world. Also, it may be very useful for the analysis and the development of the acoustical navigation systems for blind people. Firstly, the study shows that sound localization in distance and azimuth does not depend on the type of sound material. It is known that human brain is able to perceive and make decisions about sound localization when sounds are presented through the headphones. The second interesting remark is that nowadays all psychoacoustic experiments are performed in special rooms (anechoic and with reverberation) [Gröhn, 2008, etc], where the level of reverberation and noises are controlled by the experimenter, but there are no works developed in normal environments, where unexpected noises and distractions are present. In the first case all distractions are controlled; only the tiredness and the boredom are uncontrolled and strongly affect the results.

Table 4.3 Mean localization performances in percentage and standard deviation for wood, bongo and Delta material sound in anechoic environment for listener X.

	Azimuth	Distance
Wood, %	92.18	88.85
<i>sd</i>	3.13	5.91
Bongo, %	93.01	86.71
<i>sd</i>	3.98	11.53
Delta, %	93.13	93.75
<i>sd</i>	3.86	3.95

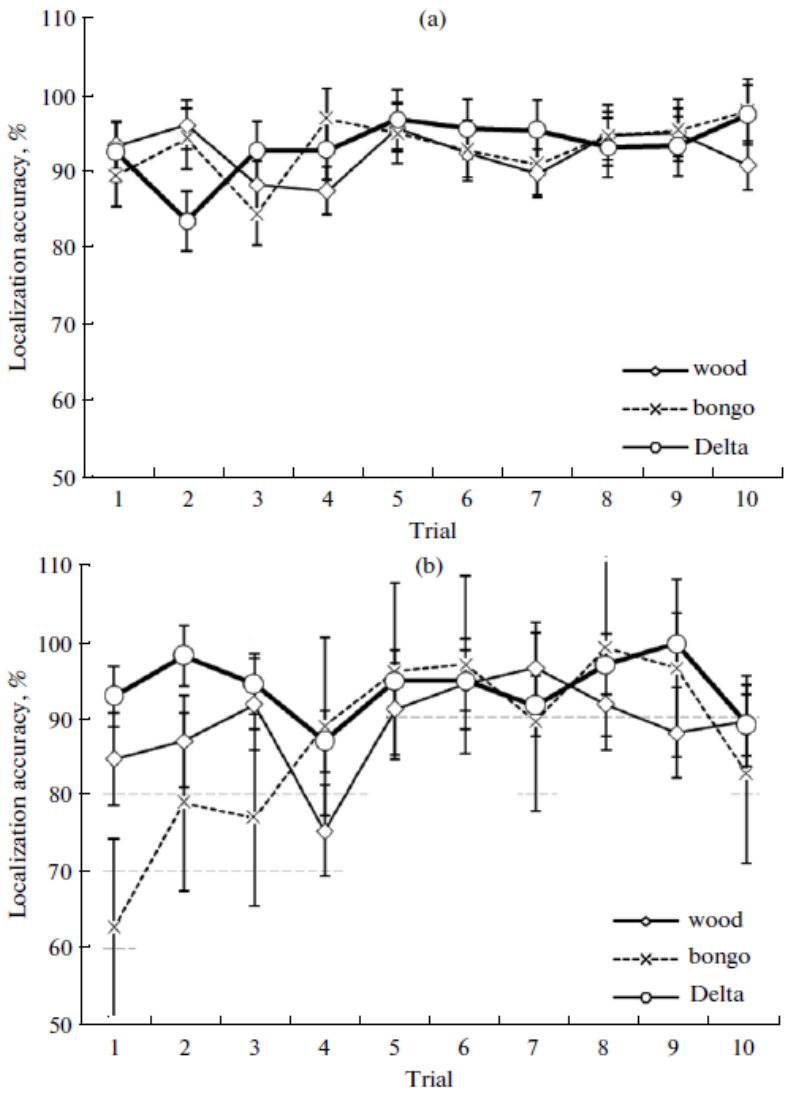


Figure 4.6 Listener X perception accuracy for anechoic sounds localization (a) in azimuth and (b) in distance. The x-axes are the trials and the y-axes represent the localization accuracy in %.

4.5 Conclusions

In the present paper, different experiments on localization of sounds coming from different materials are developed and analyzed. In order to create the spatial sensation of the sounds, as if they were coming from the real world, each sound was convolved with Head-Related Transfer Functions. These were measured for each spatial position in a frontal plane by using KEMAR dummy head. Experimental results prove that sound localization in distance and azimuth does not depend on the material sound. Also, the reverberation effect does not significantly increase the localization accuracy. Best results were obtained by the listener X, whose localization results are more constant for direction and distance.

4.6 Acknowledgments

This research was supported by Research Center in Graphic Technology from the Universidad Politécnica de Valencia.

4.7 References

- [1] D. S. Brungart and W. M. Rabinowitz, *J. Acoust. Soc. Am.* 106, 1465 (1999).
- [2] D. S. Brungart, I. Nathaniel, and W. R. Rabinowitz, *J. Acoust. Soc. Am.* 106, 1956 (1999).
- [3] H. Bruce and D. Hirsh, *J. Acoust. Soc. Am.* 31, 486 (1959).
- [4] D. I. Shore, S. E. Hall, and R. M. Klein, *J. Acoust. Soc. Am.* 103, 3730 (1998).
- [5] J. C. Kidd and J. H. Hogloben, *J. Acoust. Soc. Am.*, 116, 1116 (2004).
- [6] L. Dunai, G. P. Fajarnes, B. D. Garcia, N. O. Araque, and F. B. Simon, *Acoust. Phys.* 55, 448 (2009).

- [7] L. Dunai, G. P. Fajarnes, B. D. Garcia, and V. S. Praderas, *Acoust. Phys.* 56, 348 (2010).
- [8] M. Gröhn, *Proc. Int. Conf. on Auditory Display, Kyoto, 2002*.
- [9] E. S. Malinina and I. G. Andreeva, *Acoust. Phys.* 56, 576 (2010).
- [10] E. D. Shabalina, N. V. Shirgina, and A. V. Shanin, *Acoust. Phys.* 56, 525 (2010).
- [11] A. Pompey, M. A. Sumbatyan, and N. F. Todorov, *Acoust. Phys.* 55, 760 (2009).
- [12] R. L. Klatzky, D. K. Pai, and E. P. Krotkov, *Presence: Teleoperators and Virtual Environments* 9, 399 (2000).
- [13] M. Aramaki, M. Besson, R. Kronland-Martinet, and S. Ystad, *Proc. 5th Int. Symp. on Comp. Music Model. Retrieval (CMMR 2008)*, Copenhagen, 2008, pp. 1–8.
- [14] W. Gaver, PhD Dissertation, Univ. California, San Diego, 1988.
- [15] N. I. Durlach, A. Rigapolus, X. D. Pang, W. S. Woods, Kulkarni, H. S. Colburn, and E. M. Wenzel, *Presence: Teleoperators and Virtual Environments* 1, 251 (1992).
- [16] S. A. Gelfand, *Essentials of Audiology*, 3rd ed. (Thieme Medical Publishers, New York, 2009).
- [17] J. Jerger, *ASHA* 4, 139 (1962).
- [18] H. Mershon, W. L. Ballenger, A. D. Little, P. L. McMurtry, and J. L. Buchanan, *Perception* 18,403 (1989).
- [19] D. O. Kim, A. Moiseff, T. J. Bradley, and J. Gull, *Acta Otolaryngologica* 128, 328 (2008).
- [20] P. Zahorik, *Proc. Int. Conf. on Auditory Display, Kyoto, 2002*.

5. ARTICULO III

Ismael Lengua, Larisa Dunai, Guillermo Peris-Fajarnés, Fernando Brusola.

LOCALIZACIÓN DE SONIDOS EN ENTORNO ABIERTO POR PERSONAS CIEGAS.

Revista Acústica de la Sociedad Española de Acústica. Vol. 45. Núm 3-4
2013. I.S.S.N. 0210-3680.

Abstract

The paper describes the results of virtual sound localization via an object detector in laboratory setting and open environment with blind people.

In the tests have been involved two groups of ten blind people. Participants were aged between 25 and 60. The obstacle detection device reproduced 64 sounds at an azimuth of 64° for a distance of between 0.5 m to 5 m in a horizontal plane to the level of the eyes. Virtual sounds were reproduced through headphones.

For the experiment we used a recorded sound which varies between -110dB and -23dB with frequency between 86Hz to 18.604Hz. The sound was reproduced 44.1kHz frequency and 32 bits, and a duration of 0.04641s.

The results show that blind people are able to locate objects by virtually simulated sounds and detect them with high accuracy.

Resumen

El artículo describe los resultados obtenidos en la localización de sonidos virtuales mediante un dispositivo detector de objetos en entorno de laboratorio y en entorno abierto con personas invidentes.

En las pruebas han participado dos grupos de diez personas invidentes. Los participantes tenían edades comprendidas entre 25 y 60 años. El dispositivo detector de obstáculos reproducía 64 sonidos en un azimut de 64° para una distancia de entre 0,5 m hasta 5 m en un plano horizontal al nivel de los ojos. Los sonidos virtuales se reproducían mediante auriculares.

Para el experimento se ha utilizado un sonido grabado que varía entre -110dB y -23dB con frecuencia entre 86Hz hasta 18.604Hz. El sonido se reprodujo a una frecuencia de 44.1kHz y 32 bits, y con una duración de 0.04641s.

Los resultados nos muestran que las personas invidentes son capaces de localizar objetos mediante sonidos simulados virtualmente y detectarlos con gran precisión.

5.1 Introducción

Los seres humanos tenemos una gran capacidad de localización de una la fuente de sonido, de hecho es uno de los factores más importantes para la supervivencia humana. Esta capacidad nos ayuda a percibir, detectar peligros y diferenciar entre los cambios de las condiciones climáticas, ambientales, etc. La localización de la fuente de sonido es la "ley o regla por la cual la ubicación de un evento auditivo se relaciona con un determinado atributo o atributos de un evento de sonido "[1]. La localización de una fuente de sonido viene caracterizada por los siguientes parámetros: la diferencia de Tiempo Interaural (ITD), la Diferencia de nivel de Interaural (ILD) así como la geometría del oreja [2]. El uso de las funciones Head Related Transfer Functions (HRTF) para la localización de fuentes de Sonido Virtuales es imprescindible ya que representan las características de transferencia de la fuente de sonido en un campo libre para el oído externo del oyente.

La localización de sonidos ha sido siempre un tema de interés en acústica. Hasta el presente, se ha estudiado la localización de sonidos en dirección, llamado lateralización, para distancias cortas de un metro [1]. La mayoría de los experimentos se han producido dentro de las cámaras anecoicas o salas con reducción de ruido [3], [4]. También, el objetivo de los experimentos en localización de sonidos es la de analizar y validar la influencia de los parámetros acústicos, como la Diferencia Interaural de Tiempo [5], la Diferencia Interaural de Intensidad para las altas Frecuencias, el cono de confusión, el efecto de reverberación, etc.

En la actualidad existe poca investigación con respecto a la localización de sonidos para personas con discapacidad visual.

El presente trabajo describe los métodos y resultados de la localización de las fuentes de sonido virtual por personas ciegas mediante el uso de las funciones (HRTF).

5.2 Descripción de los equipos

El principal objetivo del proyecto es llegar a transmitir lo que un ser humano detecta cuando está caminando de forma consciente e inconsciente y que le permite poder deambular sin interceptar los objetos que va encontrando por su camino. La información que percibe el ser humano es muy amplia como puede ser la profundidad de campo, acimut, altura, etc... Toda la información hay que clasificarla previamente, no sin dificultad, para decidir cuál es la información más importante que hay transmitir para poder definir un espacio de forma unívoca mediante la representación de un plano sonoro de una escena que queramos representar. La representación del espacio mediante sonidos no es trivial y hay que definir claramente como representar los objetos de una escena. Una de las características que se aprovechar en este proyecto es la capacidad que tiene el sistema auditivo humano de localizar varias fuentes de sonido al mismo tiempo [6]. También cabe destacar la capacidad de discriminación y combinación del oído humano y su capacidad de combinarlas por clases de señales y frecuencias.

En la actualidad los sistemas de navegación existentes se basan en diferentes tecnologías de captación como son el láser, sistemas basados en estereovisión, dispositivos basados en geolocalización por satélite GPS, incluso sistemas basados en ultrasonidos. Todos estos sistemas requieren de una gran complejidad computacional del algoritmo de interpretación, lo que conlleva a requerir sistemas de procesamiento de muy alto coste. Algunos de estos sistemas transmiten la información precisa vía voz, sin llegar a ser en tiempo real con el inconveniente de confundirse con la voz humana.

La solución adoptada ha sido desarrollar un sistema cognitivo, que resuelve los principales problemas, el objetivo es transmitir al usuario la información del espacio que le rodea mediante un sistema de audio [7]. En la siguiente Figura 5.1 se muestra de forma esquemática el sistema sensorial auditivo desarrollado.

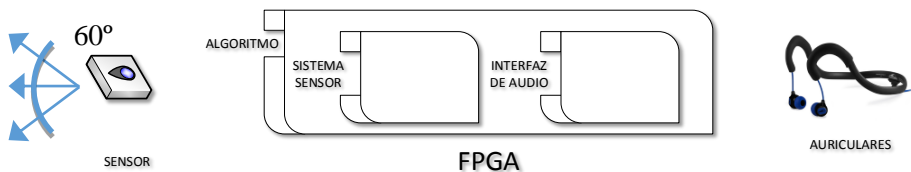


Figura 5.1 Representación esquemática del Sistema.

El dispositivo sensor está basado un CMOS láser que emite una luz láser con una definición de 64 puntos con una diferencia de un grado de separación ente cada punto, abarcando una angulación total de 64° la distancia máxima a alcanzar son 5 m. Este dispositivo nos permite obtener la distancia y azimut exacto donde está ubicado un objeto. Una vez conocida la distancia y ángulo se procesa mediante un algoritmito que atribuye a esa posición una secuencia de sonidos virtuales que permiten el posicionamiento univoco del objeto.



Figura 5.2 Imagen real con los elementos del sistema: gafas (elemento sensor), auriculares y FPGA.

5.2.1 Detección de distancias

Una vez definido todo el proceso y generados los algoritmos necesarios para el funcionamiento óptimo del sistema, el principal objetivo era integrar el dispositivo sensor de entrada en un dispositivo que fuera lo más cómodo posible y que el usuario en este caso invidente lo aceptara como un elemento más en su vida cotidiana. La solución adoptada fue integrar el dispositivo sensor compuesto por componentes ópticos, componentes electrónicos y un CMOS láser en unas gafas con apariencia similar a unas gafas de sol que la mayoría de personas invidentes suelen utilizar. El componente CMOS láser está situado justo en el centro de las gafas en la posición que correspondería a la montura justo encima de la nariz lo que proporciona una posición totalmente centrada con respecto a la lateralidad del usuario, se utilizó un polímero para simular el cristal de las gafas que no interfiere con la luz láser y tanto las pruebas realizadas para la medición de distancias con el polímero como sin polímero dieron los mismo resultados.

El sensor permite obtener las medidas de los objetos que tiene alrededor mediante los 64 pixeles que tiene el dispositivo láser que permite calcular con muy alta fiabilidad con una distancia máxima de 5 metros y una precisión de azimut de 1° dentro de una angulación limitada a 64°. El principio de medición está patentado y se basa en el momento de la medición de vuelo de la luz láser de pulso modulada Time-of-Flight (ToF). La principal ventaja de este sensor además de la precisión es que la velocidad de procesado es muy baja reduciendo el tiempo de cálculo.

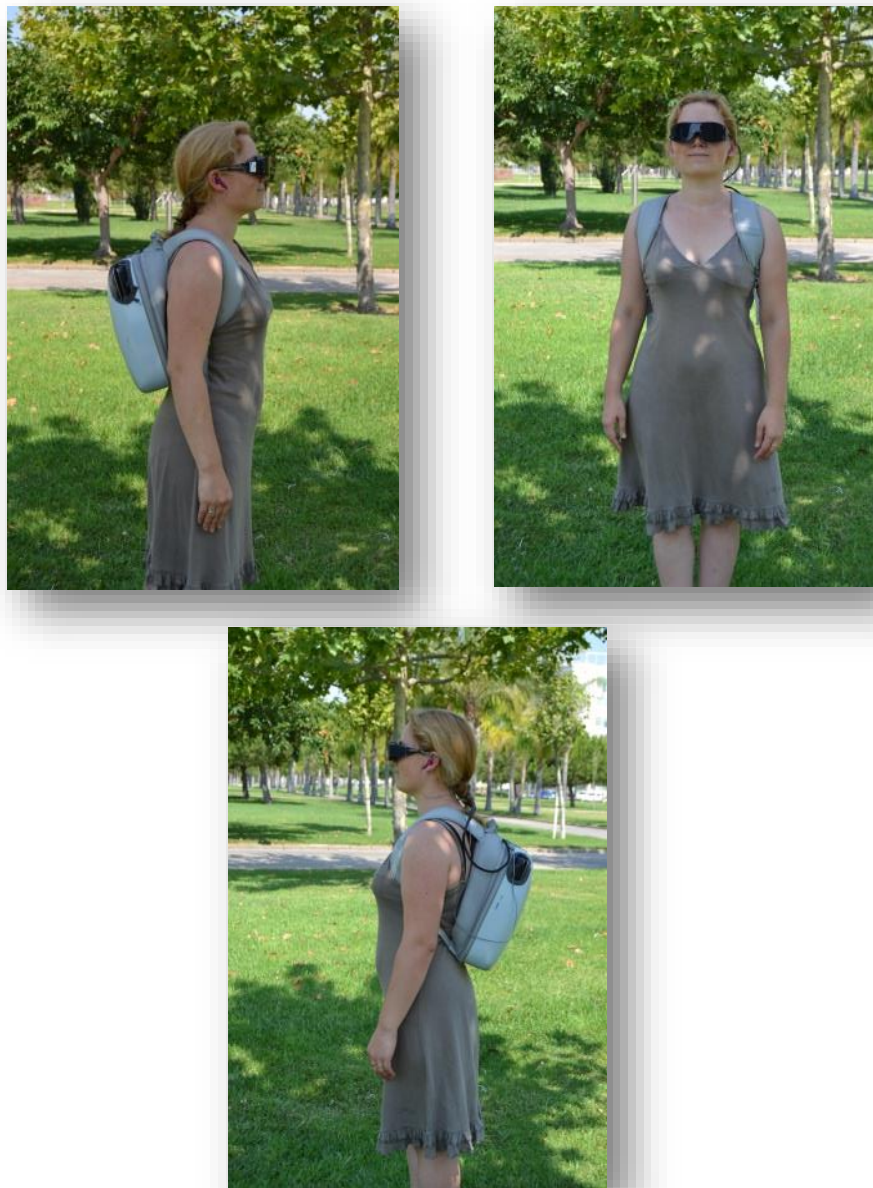


Figura 5.3 Imágenes del dispositivo en uso.

5.2.2 Sistema de audio

La información que recibe el sistema de audio es la distancia del objeto y el ángulo con respecto a un plano que abarca desde el centro del sistema sensor hasta 32° hacia la izquierda del usuario y 32° a la derecha del usuario a una distancia máxima de 5 m. Todo ello comprendido en un único plano horizontal a la altura del sistema sensor en que está integrado en las gafas en posición central.

El interfaz de audio da una respuesta al usuario en forma de clic sonoro procesados para este sistema a través de los auriculares el sonido es un clic de 2.048 muestras en frecuencia de muestreo de 44.100 Hz y de 47 ms. El sonido está convolucionado mediante Head Related Transfer Functions (HRTF) generando 64 HRTFs en acimut en cada 0,96 ° y 16 niveles en distancias de 0,5 m hasta 5 m.

Para las pruebas previas se ha utilizado Maximum Length Binary Sequence (MLBS) como fuente de sonido con el fin de medir las HRTF en una cámara anecoica.

La medición de HRTF se basa en el cálculo de la respuesta de impulso para ambos oídos mediante el filtro en el dominio de frecuencia. Para una señal de sonido $x_1(n)$ reproducida por el altavoz, la respuesta registrada se puede calcular como:

$$Y_1 = X_1 LFM, \quad (5.1)$$

donde X_1 es la representación del sonido $x_1(n)$ en el dominio de la frecuencia transformado, L es la función de transferencia del altavoz y todo el equipo de reproducción, F es la función de transferencia del espacio entre el altavoz y el oído humano y M es la transferencia función del micrófono del oído humano y todo el equipo de grabación.

Con el fin de calcular el filtro era necesario para generar un $x_2(n)$ reproducido por los auriculares tales como la respuesta registrada Y_2 para que sea igual a Y_1 . Y_2 se puede calcular:

$$Y_2 = X_2 HM, \quad (5.2)$$

donde H es la función de transferencia de los auriculares y todos los equipos de reproducción.

Entonces filtro de T se obtiene a partir de (5.4) y (5.5).

$$T = LF / H. \quad (5.3)$$

De esta manera el filtro de T se mide por las respuestas de impulso para cada oído.

La respuesta al impulso se obtiene mediante circular de correlación cruzada entre las MLBS de la entrada del sistema y la respuesta en la salida.

$$h(n) = \Omega_{sy}(n) = s(n) \Phi y(n) = \frac{1}{L+1} \sum_{k=0}^{L-1} s(k) \cdot y(n+k), \quad (5.4)$$

donde Φ representar la correlación cruzada circular.

Transformada rápida de Hadamard (FHT) se utilizó para reducir el tiempo de cálculo para la generación de HRTF:

$$h(n) = \frac{1}{(L+1)s[0]} P_2 \langle S_2 \{ H_{L+1} [S_1 (P_1 y(n))] \} \rangle, \quad (5.5)$$

donde P son las matrices de permutación, S - las matrices redimention e y H_{L+1} es las matrices Hadamard de $L+1$ grado.

En otras palabras, las señales binaurales influyeron en la generación de los sonidos virtuales. El desplazamiento de acimut de la fuente de sonido virtual se obtiene por el desplazamiento de fase interaural de las señales izquierda y derecha en términos de diferencia de tiempo entre el oído izquierdo y derecho. Cada sonido virtual corresponde a un pixel del módulo sensor y tiene una superficie de 64° .

5.3 Descripción de la prueba

Una vez verificada la fiabilidad del sistema y de los algoritmos se procedió a la realización de pruebas con personas invidentes en laboratorio. La prueba consiste en un ejercicio de detección de obstáculos, se creó un laberinto en el laboratorio mediante unas columnas formadas por cajas de cartón blando, con el objetivo de preservar de un posible impacto las personas invidentes y que no se produjera ningún daño, la columna se deshacía al mínimo impacto.

El laberinto consistía en un pasillo de 14 metros con 4 pares de columnas situados en paralelo a una distancia de 2,5 metros (Figura 5.4) en la que la persona invidente debía evitar los obstáculos y encontrar el pasillo libre [8].

El ejercicio consistía en los usuarios del sistema debían mover la cabeza a izquierda y a derecha para que el sistema les indicara donde está situado el obstáculo y así encontrar el pasillo libre.

La localización para las pruebas de navegación en interior se correspondía con un gran salón con 15 metros de longitud y una anchura de 10 metros. El intervalo de tiempo entre cada conjunto de sonidos virtuales era 153 ms.

Las pruebas fueron realizadas por 10 personas invidentes y tuvo dos etapas diferenciadas en la primera se realizaron pruebas de adaptación del usuario con el sistema de navegación, para que conociera las particularidades del sistema y controles como el control del nivel de audio para adaptarlo a sus necesidades cognitivas.

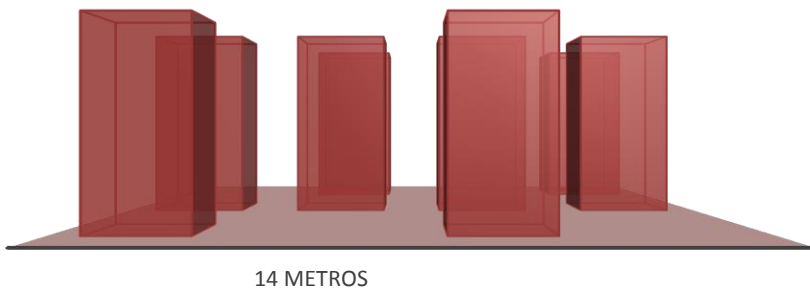


Figura 5.4 Prueba de Laberinto.

Una vez familiarizados con el sistema los usuarios invidentes procedieron a realizar la prueba dentro del laboratorio del laberinto. Las pruebas se realizaron tanto de dos formas la primera utilizando el bastón con el que están acostumbrados a navegar las personas invidentes y una segunda sin bastón.

5.4 Resultados

Lo primero que se realizó con los usuarios fue unas pruebas de adaptación al sistema para que fueran capaces de familiarizarse, todos fueron capaces en pocos segundos de percibir el origen del sonido y localizar los objetos que se probaron a modo de test. El tiempo medio de percepción del sonido era 2,32 minutos y para la detección de objetos de 4,86 minutos. El promedio de tiempo total para completar estos ejercicios fue 3,59 minutos.

Una vez familiarizados con el sistema se realizó las pruebas en laberinto. Dentro del laboratorio se tomaron datos referentes al tiempo de recorrido de las 10 personas invidentes, se realizaron dos recorridos al laberinto por usuario, obteniéndose una mejora en el resultado con respecto al tiempo de superación de la prueba de laberinto como puede observarse en el Gráfico 5.1.

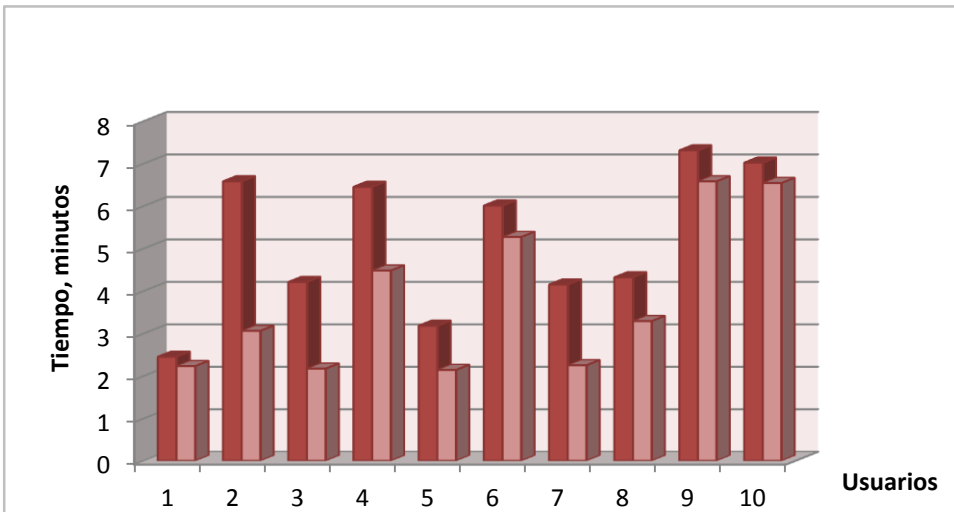


Gráfico 5.1 Tiempo de recorrido de las dos pruebas de laberinto.

Los resultados sugieren que con el sistema utilizado es posible percibir la presencia, la posición y las dimensiones del objeto detectado. Esto es lo que indica que los sonidos virtuales son una solución prometedora como una parte de la interfaz de usuario de un sistema de navegación ciego. Los sonidos acústicos virtuales tienen la ventaja adicional de que necesitan un tiempo mucho más corto que el habla convencional.

Los datos actuales aumentan y amplían los estudios anteriores, que han demostrado la utilidad de las señales perceptuales directas a la navegación. Es muy importante para el trabajo futuro para poder comparar los modelos de orientación en entornos más complejos, tales como estaciones de tren, supermercados, pasos elevados, etc... Otra línea de investigación es la localización de la fuente de sonido y señales sonoras características.

Los resultados actuales pueden tener buenas aplicaciones para el guiado en la navegación para las personas ciegas. Una ventaja del sistema es que podría integrarse con otros sistemas de navegación como el GPS y otros interfaces visuales.

5.5 Conclusiones

El sistema desarrollado ha permitido realizar las pruebas necesarias para demostrar que las personas invidentes son capaces de localizar sonidos virtuales a través de auriculares.

También hay que destacar que han sido capaces de caminar sin el bastón blanco, haciendo uso sólo de las indicaciones que le facilita el sistema teniendo más confianza en él, a medida que aumentaba el tiempo de uso. Todos los usuarios solicitaban poder llevarse el dispositivo para hacer un uso diario del sistema. El dispositivo ha demostrado que es capaz de localizar los objetos de forma unívoca y de interpretarlo de forma audible para que una persona invidente sea capaz de localizar el objeto que tiene cercano y que no es capaz de detectarlo con un bastón. El experimento también demostró que las personas invidentes tienen muy desarrollada la capacidad auditiva y una gran capacidad de adaptación y aprendizaje en el entorno sonoro.

5.6 Referencias

- [1] Blauert J. Spatial Hearing: The Psychophysics of Human Sound Localization. *Revised edn, The MIT Press*, Cambridge, MA, USA, 1997.
- [2] Brungart D.S.; Rabinowitz W. M. Auditory localization of nearby sources. Head-related transfer functions, *J. Acoust. Soc. Am.* 106(3), 1465-1479 1999.
- [3] Hartmann B. M. Localization of the sounds in rooms. *J. Acoust. Soc. Am.* 74, pp. 1380-1391
- [4] Andrew Brughera, Larisa Dunai, and William M. Hartmann. Human interaural time difference thresholds for sine tones: The high-frequency limit. *J. Acoust. Soc. Am.* Volume 133, Issue 5, pp. 2839-2855 (2013)
- [5] Fitzpatrick D. C., Kuwada S., and Batra R. Neural Sensitivity to Interaural Time Differences: Beyond the Jeffress Model. *The Journal of Neuroscience*, February 15, 2000, 20(4):1605–1615
- [6] 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.
- [7] 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.
- [8] Ismael Lengua, Larisa Dunai, Guillermo Peris-Fajarnes, Beatriz Defez. Dispositivo de navegación para personas invidentes basado en la tecnología time of flight. *Dyna*, Edición 179, pp. 33-41, Medellín, junio, 2013.

6. ARTICULO IV

Larisa Dunai, Beatriz Defez, Ismael Lengua, Guillermo Peris-Fajarnes.

3D CMOS SENSOR BASED ACOUSTIC OBJECT DETECTION AND NAVIGATION SYSTEM FOR BLIND PEOPLE.

IECON 2012 - 38th Annual Conference on IEEE Industrial Electronics Society. 4208-4215 Montreal, QC Date of Conference: 25-28 Oct. 2012, ISBN: 978-1-4673-2419-9.

Abstract

The paper presents a new wearable Cognitive Aid System for Blind People (CASBlIP). The prototype device was developed as an obstacle detector, orientation and navigation Electronic Travel Aid (ETA) for blind people. The device provides a binaural acoustic image representation. The environmental information acquisition system is based on an array of 1x64 CMOS Time-of-Flight sensors. Through stereoscopic (binaural) acoustic sounds the device relays the surrounding near and far environment. Experimental results demonstrate that blind users are able to detect obstacles and navigate through unknown and known environments safely and confidently. CASBlIP works accurately in range of 15m in distance and 64° in azimuth, providing significant advantages in comparison with currently existing ETA systems.

6.1 Introduction

There are over 314 million of blind and partially impaired people in the world. Over 45 million of them are totally blind [1]. Besides independence and communication limitation, the main problem for blind people is their mobility. Mobility has been defined by Foulke [2] as “the ability to travel safely, comfortably, gracefully and independently through the environment”.

Since early decades, together with the sensors development, many efforts have been made to develop new and sophisticated Electronic Travel Aids (ETAs) able to perceive and represent the surrounding environment. The idea

underlying the development of such devices was to overcome human sense limitations, such as blindness. These devices would help blind people to perceive their surrounding environment through tactile, vibrotactile, vocal or acoustic senses.

Traditionally, three classes of ETAs are known: obstacle detectors, environmental sensors and orientation and navigation aids. The first two classes make use of ultrasonic, laser and artificial vision to measure the distance to the nearest objects or receptors. The third class uses global positioning systems (GPS) for route prediction and orientation task.

The first remarkable obstacle detector ETA is the Lindsay Russell Pathsounder [3], [4]. Also, the Mowat Sonar Sensor is one of the simplest and most popular [5]. Both devices use ultrasonic transducers which measure the distance to the nearest object, providing tactile or vibrations feedback [6], [7]. This subgroup of ultrasonic based obstacle detector ETAs also includes other devices such as Nottingham Obstacle Detector, Sonic Torch, Sonic Pathfinder, Polaron™ Sensory 6, WalkMate [8], [6], [9], [10]. The existing obstacle detectors ETA devices are able to detect the obstacles in a range covering up to 15 feet and 15° in azimuth.

On the other hand, the most popular laser obstacle detector ETA is the Laser Cane [8], [11]. Laser Cane is based on laser technology which is implemented into the cane, making use of the triangulation principle. Its photo detectors receive the light beams which, after collision with the obstacle surface, are reflected. This enables the device to calculate the distance up to the obstacle. The detection range is up to 4m. Through acoustic signals, the blind user is informed about detected object.

The second class, environmental sensors, includes devices that try to go further than obstacle detectors devices [6], [11], providing additional features. Bat K Sonar Cane [12], Ultracane, KASPA, Trisensor, Miniguide, Sonic Torch, Sonicguide and Navbelt [13] belong to this class. The most influential ETA from environmental sensors is the Binaural Sonar Electronic Travel Aid also known as Sonicguide [14]. Sonicguide is the improved Sonic Torch device, developed by Kay in 1959 [8], [15]. The device incorporates two acoustic channels which help blind users not only to detect the objects but also to recognize and count them in range up to 4m in distance and 55° in azimuth [16], [14]. On the other hand, vOICE device is one of the latest ETA systems using stereovision technology for environmental image acquisition. The device, via acoustic signals and a tactile interface, informs the users about their surrounding environment [17], [18].

The third class of ETAs corresponds to the orientation and navigation aids; the main objective of these devices is the familiarization with the environments as well as the navigation [19], [20], [21]. The pioneer orientation device is Talking Signs [22]. The device, by means of an infrared beam pattern, controls the range and coverage of each sign in a range up to 30m in outdoor environments. SONA – Sonic Orientation Navigation Aid is based on a similar idea as Talking Signs [23]. The system was developed at Georgia Institute of Technology in 1979. Other devices such as Verbal Landmark [24], Easy Walker [11], Pilot Light, etc...belong to this class too. Also, the Global Positioning Systems (GPS) [10] are included in this group, [25]; among them, one of the best known is the GPS ETA, developed at the University of California [26], and known as Personal Guidance System [27]. This device is based on radio signals from satellites, which provide real information about every point of the Earth surface [28], [29]. Through synthetic speech [30] or tactile [31] the user gets informed about his location. The system range is 20m. Within this group we can also find the Mobility for Blind and Elderly People Interacting with Computers (MoBIC) device [32], Makino [27], Electronic Guide [27], GPS Braille Note [33], Geotact [34] Easy Walker [35], Loadstone, Trekker [36] or Tormes [37], Tanea prototype [38], [39], System for Wearable Audio Navigation (SWAN) [40], [41], Tyflos [42], [43], [44], [45], [46]. All these devices provide artificial speech, acoustic or tactile feedback to the user.

The aim of this paper is to propose a new system able not only to enable mere object detection, but also to provide orientation and navigation aid. This system is specially suited for blind and partially blind people. It provides significant advantages against other conventional ETA systems, such as wider detection range, possibility of use both in indoor and outdoor environments as well as a simple and easy utilization. The paper is structured as follows: Section II provides a general description of the developed device, listing its main advantages and drawbacks. Section III details the technical implementation of the sensor module whereas Section IV describes the acoustic module. In Section V, the experimental results are presented and analyzed. Finally, Section VI shows the conclusions of the work.

6.2 Description of the Acoustic System

Advances in sensor technology and electronics allow providing environmental information to visually impaired people. In the present section, the developed object detector and navigation Electronic Travel Aid is described.

In general terms, the CASBlIP system detects the objects in the frontal view of the user and converts the information regarding object location into binaural virtual sounds which are delivered to the user via stereophonic headphones. The object detection is achieved by emitting laser light beams which are reflected after colliding with the surrounding objects (see Figure 6.1). These light beams are received by the sensor system of the device and converted into sounds by means of its internal software.

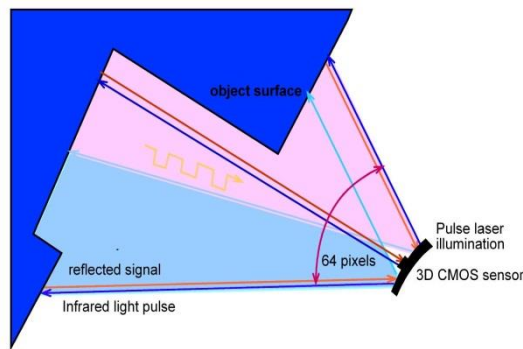


Figure 6.1 Measurement principle of the Time of Flight (TOF) method.

As it will be later detailed, the object detection system of the device is based on an improved 1×64 3D CMOS image sensor, originally developed for pedestrian recognition by cars [47], [48]. This sensor is implemented into a pair of glasses which are also equipped with the laser electronics for light beam emission. Besides, the Field Programmable Gate Array (FPGA), in charge of information processing and conversion into sounds, is implemented in an additional module of the device. This FPGA contains functions regarding sensor control, data acquisition, distance calculation, application software, sounds files, sound reproducing system, power supply and connectors between the parts of the device. As a conclusion then, the device is based on two main parts; sensor module, which includes the glasses and part

of FPGA (for laser beam emission and reception) and the acoustic module based on the part of the FPGA in charge of information processing and conversion into sounds and on the stereo headphones (to deliver the sounds to the user).

The aim of the Acoustic System is not only to inform the user about the object presence in the range of the device, but also to inform him about object direction and distance. The device operates in an on-line mode. This means that it delivers environmental information in real-time. Moreover, the device works properly both in known and unknown environments, and both in simple and in complex situations. Acoustic sounds perceived via headphones by the human brain create a map of the detected object, leaving the cognition task to the user. By hearing virtual sounds, the user is in charge to decide whether to avoid the obstacles in his direction or not. In this context, the functions of the Acoustic System are the detection and information tasks, whereas the navigation and route planning tasks are carried out by the user.

Although there are many Electronic Travel Aids that implement acoustic sounds, such some of those commented in Section I, there are significant differences between the developed acoustic system and other acoustic ETAs.

- The acoustic system enables the user to decide route and path planning. The device detects all objects appearing in the area of view. At the same time it informs the user about their presence and the user is free to decide to decide his way.
- The device provides a full environmental image due to the 1×64 pixels 3D CMOS image sensor which covers 60° in azimuth and a distance range between 0,5m and 5m. The device simultaneously provides 64 sounds covering the whole detected environment.
- The device perceives the object depth, since the different distances are represented by specific sounds.
- Virtual acoustic sounds are binaural sounds; the human brain easily perceives them as sounds provided by the environmental objects.
- The device provides environmental information in real-time. Sensor and Acoustic modules operate so fast that the device does not require supplementary time to interpret the environmental information.
- No false noises; when there is no object in the area of view of the system, the device is silent, and so no sounds are reproduced by the system.
- The system works on darkness and sunny days, cold and hot weather, because the 3D CMOS sensor works with infrared laser impulses.

- Acoustic sounds are sounds which do not represent any impact material, also do not disturb the user when the device is used during a long time.
- The environmental acquisition module of the device is placed at the human eyes level, which allows the user to perceive the surrounding from the head level. Since it is a head mounted device, the user is able to move the head and scan the environment at the direction of view of the user.
- Because of the binaural cues of the acoustic sounds, the sounds do not overlap when there is more than one object in the area of view.
- With the device the user is able to perceive the object shape, i.e., if it has a plane, convex or concave surface.
- The used sound does not interfere with external noises such as traffic lights, human noises, etc.

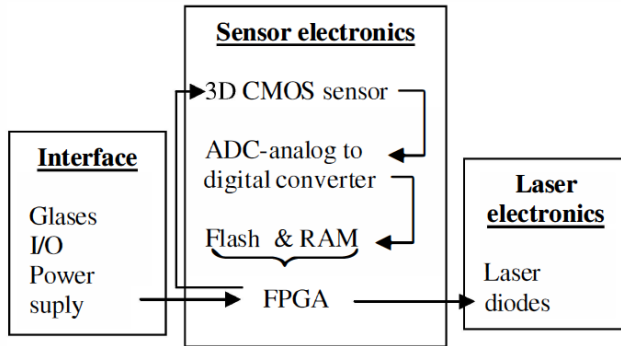


Figure 6.2 Multiple Double Short-Time Integrated measurement principle.

6.3 Implementation of the sensor module

The environmental image acquisition device or *sensor module* consists of a 4x64 3D CMOS image sensor based on 1x64 photodetector pulses.

The design of the sensor module is based on three parts: interface, sensor electronics and laser electronics. Fig 2 shows the sensor system design, where the laser pulse emitters are placed at the left and right side of the 3D CMOS sensor close together on a pair of glasses. The processing unit of the sensor

module is implemented into FPGA, which is in charge of sensor control, data acquisition, and distance calculation and application software.

Sensor module transmits and receives identical 1x64 pixels laser pulses with refreshing rate of 25fps in a horizontal line at the user eye level, where the distance range is strictly dependent of the sensor module. Sensor module range for object detection is between 0,5m and 5m in distance and up to 60° in azimuth. 3D CMOS sensor generates pulses in the range of 20 to 200ns, the rise and fall timing range is 10ns, maximum shutter speed is 30ns, pulse repetition frequency is up to 20 kHz and the laser pulse peak power is up to 2kW. The wavelength is 905 nm in the near-infrared laser pulse (NIR) range; therefore the laser light becomes invisible to the human eye. Maximum frame rate of the image sensor is 19500 fps for a synchronic clock of 66 MHz, each pixel clock is of 5MHz, and output swing is 1,9V. The generated pulses are strongly rectangular pulses. The pixel pitch is 130µm in horizontal direction resulting in a sensitive area of 64x130µm = 8,32mm². The chip is completely controlled by the FPGA.

Laser pulses generated by the NIR laser diodes and defocused by diffractive optics are emitted to the environment illuminating entire field of view [49]. When an object appears in the laser range, the pulses are reflected to the camera lenses of the 3D CMOS image sensor calculating the distance between the object surface and the sensor system for every 1x64 pixels simultaneously (see Figure 6.1).

The time delay τ_{TOF} of the laser light between the sensor and the object for a distance d can be calculated as.

$$\tau_{TOF} = \frac{2 * d}{c} \quad (6.1)$$

The Time-of-Flight (TOF) measuring method is based on the fact that the emitted light pulse is not measured directly, but through accumulation of charges on a sensor chip. This procedure is equivalent to the Continuous wave distance measuring method when using a short single pulse [47]: Camera lenses transmit a part of the reflected laser pulse to the 3D CMOS sensor chip surface, which operates through its light sensitive pixels as if they were electronic short-time windowed integrators. Then the distance between object surface and 3D CMOS sensor is calculated as:

$$d_x = c \frac{\tau_{TOF}}{2} = \frac{v_c}{2} (T_1 - T_w \frac{U_1}{U_2}), \quad (6.2)$$

where T_1 represents the short integration time, T_w is the pulse width of the linear laser, U_1 is the measured sensor signal at integration time T_1 and U_2 is the sensor signal measured for integration time T_2 , d is measured distance, v_c is light propagation speed. The propagation time can be calculated by computing the quotient of the two integrated shutter intensities U_1/U_2 .

Figure 6.1 shows the measurement principle of indirect TOF method [47] used in the development of the Acoustic System.

The travel time between the emission and reception of the laser pulse at pixel x depends on the travel distance, and can be calculated as:

$$T_{Travel,x} = \frac{2v_c}{d_x} . \quad (6.3)$$

Due to the near-infrared laser pulses, which have the ability to illuminate the entire scene simultaneously avoiding any sequential scanning, Time-of-Flight system process images in real time. In order to reduce the required laser power and to increase the measurement accuracy, multiple double short-time integration algorithms (MDSI) were used.

6.4 Implementation of the Acoustic Module

Whereas the sensor module provides to the acoustic module the linear image of the surrounding environment, the acoustic module, is in charge of transmitting this information to the blind user, by using virtual acoustic sounds. The principle of the acoustic module is to assign an acoustic sound to each one of the 64 pixels, for different distances. The acoustic sounds will be reproduced through the headphones from the position of the detected object, always that the sensor will send to the acoustic module distance values. The sound module contains a bank of acoustic sounds previously generated for a spatial area between 0,5m and 5m, for 64 image pixels, where one pixel has

the coordinate of $0,94^\circ$ in azimuth occupying the 30° at the left ear and 30° at the right ear of the human head. A delta sound of 2040 samples at a frequency of 44,1kHz was used to generate acoustic information of the environment. In order to define the distances, 16 planes in distance were generated from 0,5m increasing exponentially up to 5m. Refreshing rate of the sounds is 2fps. 16 Mb memories are needed to process acoustic module. Distance displacement is strongly dependent on the sound intensity and on the pitch. At nearer distances, the sound is stronger than at farther distances. The more the distance increases, the more the sound intensity decreases. Virtual sounds were obtained by convolving an acoustic sound with non-individual Head-Related Transfer Functions (HRTF) previously measured using KEMAR manikin described below. Figure 6.3 shows measurement model of Head-Related Transfer Function.

The working principle of the acoustic module is similar to ‘read and play’. This is, the acoustic module reads the output data, which represent coordinates in distances and in azimuth, from the sensor module and plays the sound with the same coordinates. The time interval between sounds is 8ms when there are sounds playing. When there are no sounds, after 5ms the sound module recalls the sensor module.

To measure the HRTFs a Maximum Length Binary Sequence (MLBS) was used to generate the sound source. HRTF measurement is based on the calculation of the impulse response for both ears by using the filter in frequency domain.

For a sound signal $x_1(n)$ reproduced by the speakerphone, the registered response can be calculated as:

$$Y_1 = X_1 LFM, \quad (6.4)$$

where X_1 is the representation of the sound $x_1(n)$ in the transformed frequency domain, L is the transfer function of the speakerphone and all reproduction equipment, F is the transfer function of the space between the speakerphone and human ear and M is the transfer function of the microphone from the human ear and all recording equipment.

In order to calculate the filter, it was necessary to generate a $x_2(n)$ - sound reproduced by headphones- so that the registered response Y_2 would be equal to Y_1 . Y_2 can be calculated as:

$$Y_2 = X_2HM, \tag{6.5}$$

where H is transfer function of the headphone and all reproduction equipment. Then, the filter T is obtained from (5.4) and (5.5).

$$T = LF/H, \tag{6.6}$$

In this way, the filter T is measured for the impulse responses for each ear. The impulse response is obtained by circular cross-correlation between the MLBS from the system input and the response at the output.

$$h(n) = \Omega_{sy}(n) = s(n)\Phi y(n) = \frac{1}{L+1} \sum_{k=0}^{L-1} s(k) \cdot y(n+k) \tag{6.7}$$

Where Φ represents the circular cross correlation.

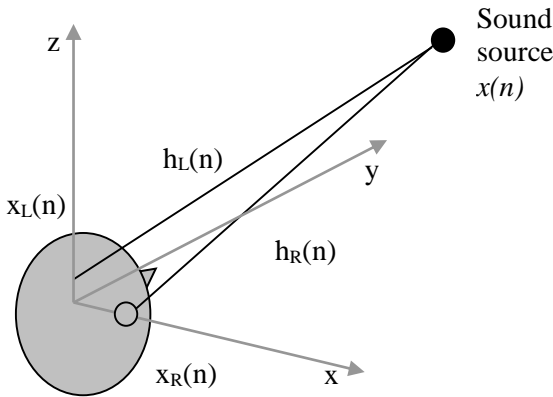


Figure 6.3 Head-Related Transfer Function measurement model.

Direct implementation of (6.7), for large sequences becomes impossible due to its computational time. Then, the convolution method was used:

$$a(n) \Phi b(n) = (1/(L+1))a(-n)*b(n), \quad (6.8)$$

And even then the computational time is great due to that the Fast Fourier Transform used is of $2k-1$ length. In that case, the Fast Hadamard Transform (FHT) was used to reduce twice the computational time for HRTF generation:

$$h(n) = \frac{1}{(L+1)s[0]} P_2 \langle S_2 \{ H_{L+1} [S_1 (P_1 y(n))] \} \rangle, \quad (6.9)$$

where P are the permutation matrices, S are the redimension matrices and HL+1 are the Hadamard matrices of L+1 degree.

In conclusion, binaural cues had an influence on the generation of the virtual sounds. The azimuth displacement of the virtual sound source was obtained by the interaural phase shift of the left and right signals in terms of time difference between the left and right ear. Each virtual sound corresponds to one pixel from the sensor module and covers an area of 60° .

6.5 Experimental Results

In this section, the tests with the CASBlIP System are described. Twenty blind users took part in the experiments. These experiments were developed both in laboratory and in outdoor environments not known ‘a priori’, under the supervision of blind people trainers and engineers. In order to avoid unnecessary displacements, two groups, group A (users from Germany)

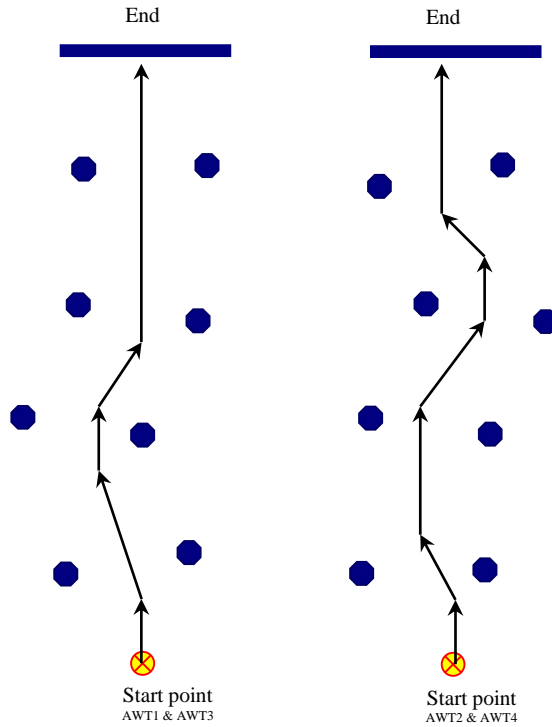


Figure 6.4 Laboratory experimental scenario.

and group B (users from Italy) simultaneously tested the device at their respective countries. For the experimental task, two identical devices were used, one for each group. Before proceeding to the experimental task of object detection and navigation, subjects were trained during a month. The aim of this training was to learn to externalize the sounds listened through the headphones as well as to comprehend how to relate one spatial coordinate from the near environment to one specific virtual acoustic sound listened through the headphones, noticing the difference between two different located sounds. Seven exercises with different levels of complexity were developed. These exercises included from single object detection to the detection of several objects, doors and walls. During the tests, users were placed at distance of 2m from the object and they were asked how they perceived the

sounds. When the users were prepared, they were allowed to move towards the object and back, going around the object. The idea was to learn how to localize spatial sounds. It is important to remark that, for blind people, each spatial sound represents an obstacle which must be avoided. Training period depended of the user ability to localize the sounds. After the training period, blind users proceeded to the laboratory and outdoor experimental trials.

The experiments aimed to test the Acoustic System as obstacle detector and navigation mobility aid for blind people. The tests consisted of two tasks: navigation through a path with 4 pairs of soft obstacles in a controlled environment and navigation in an open uncontrolled environment. In the first experiment, subjects

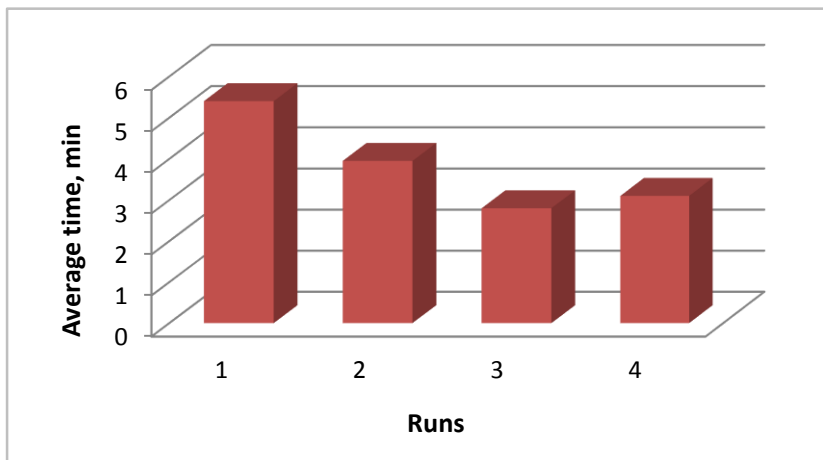


Figure 6.5 Average Absolute Walking Time (AWT) for laboratory experiment.

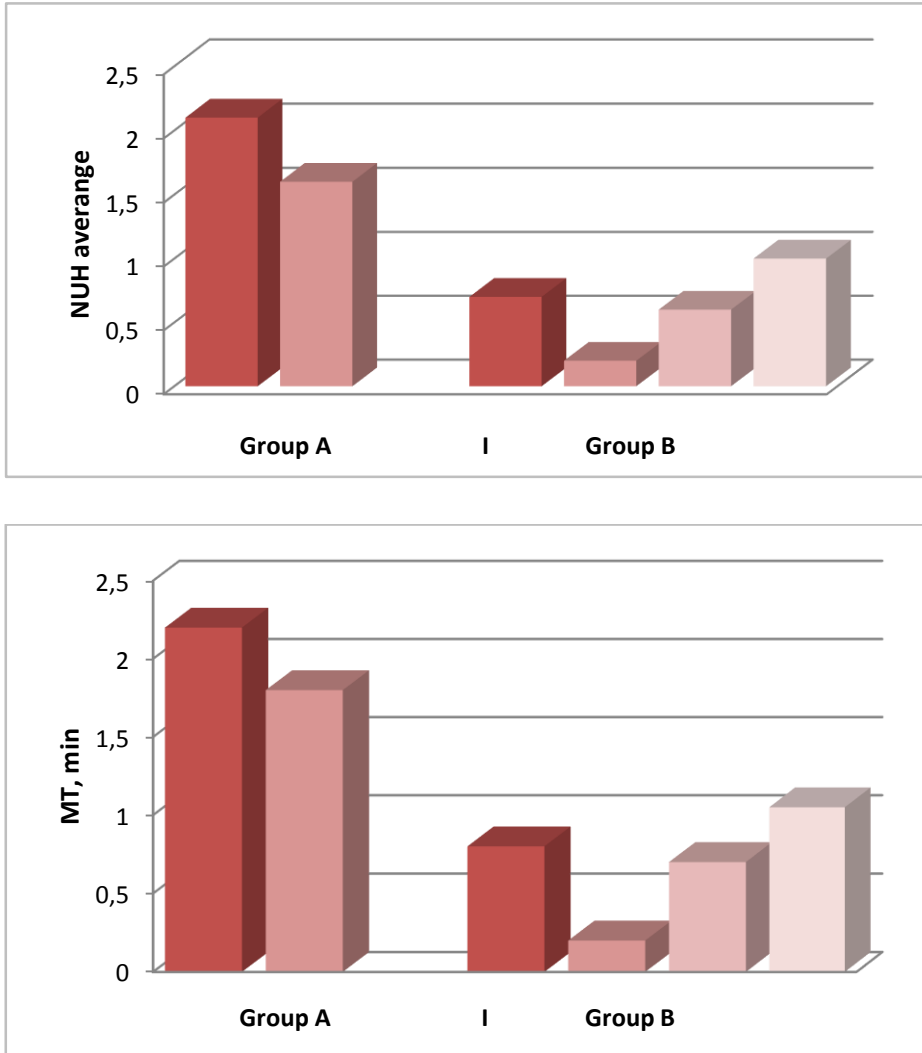


Figure 6.6 Mean absolute walking times from Mobility Test I and mean number of hits. The ten clusters of bars show mean results for the ten subjects from the group A as well as the mean results from the four runs of the Group B. The cluster of bars from the second graph shows the average Number of Hits for both groups for the previous carried out runs.

walked through a 14m artificial laboratory path limited by soft carton boxes (see Figure 6.4). All obstacles were identical in weight and height. After the device had been connected and prepared, the users were faced to the scenario. The columns were placed in an asymmetric order, separated horizontally by 2m and at a distance of 2,5m between the last pair of columns and the wall. In the laboratory experiment all obstacles remained static. The objective of the experiment was to navigate through the path avoiding all the obstacles. Two routes were prepared for the experiment. The second route represents the opposite direction to the first, i.e., the way back direction of the first route. In the experiment, the absolute walking time (AWT), number of hits (NUH) and number of corrections (NOC) were measured. Blind subjects were able to perceive object thickness and texture. Also the device can reliably detect objects with small diameters due to its 64 pixels which scan only 60° in azimuth.

The data are presented for a clear inspection in Figure 6.5. The runs 1 and 2 are the runs for the group A and the runs 3 and 4 are the runs for the group B. Note that after the training exercise, the participants performed the subsequent experiments without any white cane, trainer instructions or support. The value of time in minutes in which the subject reported the end of the trajectory is plotted in the y-axis. Figure 6.5 shows that the group A required a longer time to complete the experiment. This is due to the fact that some subjects which had difficulties on sound externalization and orientation took part in the experiment. The difficulties were due to the time of blindness; for example, in the group A, one of the subjects is blind from born and he never used a navigation device for blind people. This user find difficulties in learning that the sounds were representing something that he had never seen, or that the sounds listened via headphones were not in his head, but coming from the environment.

The goal of the second experiment, named Mobility Test I, was to test perceptual authenticity of the navigation, by evaluating if blind users were able to detect the objects and to navigate through a 29m scenario in an outdoor environment.

Table 6.1 Mobility test results for twenty subjects representing the Absolute Walking TIME (AWT), the Number of Hits (NH) and the Number of Corrections (NOC).

Mean time, min	Group A	Group B
AWT-0	1,00	-
AWT-1	9,02	8,27
AWT-11	7,31	6,46
AWT-2		6,49
AWT-21		5,55
NUH-1	2,80	0,70
NOC-1	1,40	0,20
NUH-11	2,00	0,60
NOC-11	1,20	1,00

In the evaluation of object detection and spatial sound localization in an open environment, there is always the problem that the sounds are not be perceived exactly as in the laboratory since, in the laboratory, the user is protected from the external noises, such as traffic noises, human, animal and bird’s speech, etc. Also, blind users require longer times to take decisions about the perceived objects and the environment.

Blind user tasks were to localize the objects and to judge the differences between the perceived objects, to analyze the environment, to select the free path and walk through the scenario. External noises had a slight influence in the object detection performance.

Four different scenarios were used for this test. The group A, from Germany, used two external scenarios. Both scenarios were located in the patio of the School for Blind People of Berlin (ABSV). 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 intersected the user direction of view. Because no objects appeared in the system direction of view, no signals were sent by the system to the user. Thus, the user knew that in an area of 5 meters, there were no obstacles.

Also, the user knew that when he was listening a sound, then there were objects are in his area of view. Despite the total silence of the system the user knew that this silence did not mean that the system was broken, so he must

not worry. The group B used also two different scenarios. The tests were carried out in the IFC Institute courtyard in Italy.

Group A carried out two sessions of this test (in two different days) whereas the group B carried out four sessions. The location of the soft obstacles was modified except on the real obstacles such as building wall, columns and bench.

This time, the walking time, number of hits and number of corrections were compared between the two groups (Figure 6.6). In average, the group A spent 90,19 minutes for the first run and 75,09 minutes for the second run. 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. Table I confirms that a slight improvement was detected for the repeated run. Also, we can observe that all subjects showed a very similar pattern in the first run, except on four subjects, which had great difficulties in orientation. Three of them, apart from requiring more time to navigate through the trajectory, were less accurate in the navigation. These subjects had difficulties in object detection making more hits and confusions and needing more corrections. According to the results from Figure 6.5 and Figure 6.6, navigation time is affected mainly by the lack of training.

6.6 Conclusions

In the paper, a new wearable cognitive real-time Acoustical System device is described, which was developed as obstacle detector, orientation and navigation Electronic Travel Aid for blind people. An array of 1x64 CMOS Time-of-Flight sensors scan the environment and through binaural acoustic sounds the device relays the surrounding near and far environment. The prototype device aims to be a portable device and constructed by commercially available components.

x

Finally, it must be also said that future improvements are possible in order to obtain more complete environment information:

- Improvement of the image acquisition techniques by adding more arrays of 1x64 CMOS image sensor, or using stereovision instead CMOS image sensor.

- Improvement of the acoustic module adding sounds in elevation.

6.7 Acknowledgment

The authors wish to make the following acknowledgments to the participants of CASBlIP project and to all blind people for their active participation on the system development, tests and evaluation. The project was developed within the 6th European Frame Programme, Cognitive Systems and Robotics.

6.8 References

- [1] World Blind Union, “White cane safety day,” World Blind Union Press release, October, Canada, 2009.
- [2] E. Foulke, “The perceptual basis for mobility,” AFB Res. Bull, vol. 23, pp. 1, 1971
- [3] L. Russell, “Travel Path Sounder Proceedings,” Rotterdam Mobility Research Conference. N.Y: American Foundation for the Blind, 1965
- [4] R.W. Mann, “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, 1970
- [5] D. L. Morrisette, G.L. Goddrich, J.J. Henesey, “A follow-up-study of the Mowat sensors applications, frequency of use and maintenance reliability,” J. Vis. Impairment Blindness, Vol. 75, 6, pp. 244-247, 1981
- [6] H. A. M. Freiburger, “Mobility Aids for the Blind,” Bulletin of Prosthetic Research, pp. 73-78, 1974
- [7] Book, “Electronic travel aids: new directions for research,” National Research Council, Chap. 6, pp. 67-90, 1986
- [8] L. W. Farmer, “Mobility devices,” Bulletin of Prosthetic Research, pp-47-118, 1978
- [9] N. Debnath, J. B. Tangiah, S. Pararasaingam, A. S. A. Kader, “A mobility aid for the blind with discrete distance indicator and hanging object detection,” IEEE TENCON Vol.D, pp. 664-667, 2004

- [10] L. Farmer, D. Smith, "Adaptive technology," Book chapter, Foundations on Orientation and Mobility, Eds. 2, pp.231-260, 1997
- [11] B. Ando, "Electronic Sensory systems for the Visually Impaired," IEEE Instrumentation & Management Magazine, pp. 62-67, 2003
- [12] B. S. Hoyle, "The Bathcane – mobility aid for the vision impaired and the blind," IEE Symposium on Assistive Technology, pp. 18-22, 2003
- [13] S. Shoval, J. Borenstein, Y. Koren, "The Navbelt-A computerized Travel Aid for the blind based on mobile robotics technology," IEEE Transactions on Biomedical Engineering, vol. 45, no. 11, November 1998, pp.1376-1386
- [14] R. Kuc, "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, 2002
- [15] L. Kay, "An ultrasonic sensing probe as an aid to the blind," Ultrasonics, Vol. 2, No. 2, pp. 53-59, 1964
- [16] L. Farmer, D. Smith, "Adaptive technology," Book chapter, Foundations on Orientation and Mobility, Eds. 2, pp.231-260, 1997
- [17] P. B. L. Meijer, "An experimental system for auditory image representations," IEEE Transaction Biomedical Engineering, BME-39, (2), pp. 112-121, 1992
- [18] P. B. L. Meijer, "A modular synthetic vision and navigation system for the totally blind," World Congress Proposals, 2005
- [19] J. Brabyn, "New Developments in Mobility and Orientation Aids for the Blind," IEEE Transactions on Medical Engineering, Vol. BME-29, NO 4, pp. 285-289, 1982
- [20] J. Brabyn, "A review of mobility aids and means of assessment," In D. H. Warren & E. R. Strelow (Eds) Electronic Spatial Sensing for the Blind, Martinus Nijhoff, Boston, pp.13-27, 1985
- [21] J. Brabyn, "Technology as a support system for orientation and mobility," American Rehabilitation, 1997
- [22] J. Brabyn, "New Developments in Mobility and Orientation Aids for the Blind," IEEE Transactions on Medical Engineering, Vol. BME-29, NO 4, pp. 285-289, 1982

- [23] R. Kuc, "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, 2002
- [24] B. Bentzen, P. Mitchell, "Audible signage as a wayfinding aid: comparison of verbal landmark and talking signs. Draft Report to the American Council of the Blind, Sept. 1993
- [25] C. C. Collins, "On mobility aids for the blind," in *Electronic spatial sensing for the blind*, D. H. Warren & E. R. Strelow, Eds., Dordrecht, Martinus Nijhoff, pp. 35-64, 1985
- [26] J. Loomis, R. Golledge, "Personal guidance system using GPS, GIS, and VR technologies," *Proceedings, CSUN Conference on Virtual Reality and Person with Disabilities*, San Francisco, June 17-18, 2003
- [27] J. M. Loomis, R. G. Golledge, & R. L. Klatzky, "GPS-based navigation systems for the visually impaired," In. W. Barfield & T. Caudell, Eds, *Fundamentals of wearable computers and augmented reality*, pp. 429-446, Mahwah, NJ: Lawrence Erlbaum Associates, 2001
- [28] J. M. Loomis, R. G. Golledge, R. L. Klatzky, & J. R. Marston *Assisting wayfinding in visually impaired travelers*. In G. Allen (Ed.), *Applied spatial cognition: From research to cognitive technology*. Mahwah, N. J.: Lawrence Erlbaum Associates, 2006
- [29] J. M. Loomis, R. G. Golledge, & R. L. Klatzky, "Navigation system for the blind: Auditory display modes and guidance," *Presence: Teleoperators and Virtual Environments*, 7, pp.193-203, 1998
- [30] J. M. Loomis, J. R. Marston, R. G. Golledge, & R. L. Klatzky, "Personal guidance system for people with visual impairment: A comparison of spatial displays for route guidance," *Journal of Visual Impairment & Blindness*, 99, pp. 219-232, 2005
- [31] R. L. Klatzky, J. R. Marston, N. A. Guidice, R. G. Golledge, J. M. Loomis, "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
- [32] H. Petrie, V. Johnson, T. Strothotte, A. Raab, R. Michel, L. Reichert, A. Schalt, *MoBIC: an aid to increase the independent mobility of blind*

- travellers,” *British Journal of Visual Impairment*, Vol. 15, No. 2, pp. 63-66, 1997
- [33] HumanWare “Braille Note GPS”, www.humanware.com
- [34] R. Fancy, R. Leroux, A. Jucha, “Electronic travel aids and electronic orientation aids for blind people: Technical, rehabilitation and everyday life points of view,” CVHI Conference, 2006
- [35] Action for Blind People. Easy Walk – a new system for people with sight loss. <http://www.actionforblindpeople.org.uk>
- [36] Y. Lagace, “Trekker GPS system creates handheld personal guide with HP iPAQ Pocket PC,” HumanWare, HP notice, 2005 April
- [37] T. M. Morales, A. M. Berrocal, “El GPS como sistema de ayuda a la orientacion de personas ciegas,” IICCV-INTEREVISUAL Autonomia Personal, 2005, October
- [38] A. Hub, S. Kombrink, K. Bosse and T. Ertl, “TANIA – a tactile-acoustical navigation and information assistant,” CSUN Conference. Conference Proceedings of the California State University, Northridge Center on Disabilities’ 22nd Annual International Technology and Persons with Disabilities Conference, Los Angeles, CA, USA, 2007, March 19-24
- [39] A. Hub, D. 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, pp.147-152, 2004
- [40] J. Wilson, B. N. Walker, J. Lindsay, C. Cambias, F. Dellaert, “SWAN: system for wearable audio navigation,” 11th IEEE International Symposium on Wearable Computers (ISWC 07), pp. 91-98, 2007, 11-13 Oct.
- [41] B. N. Walker, J Lindsay, “Using virtual environments to prototype auditory navigation display,” *Asst. Technology*, 17, pp. 72-81, 2005
- [42] N. Bourbakis, D. Kavraki, “Intelligent assistants for handicapped people's independence: case study,” *IEEE International Joint Symposia on Intelligence and Systems*, Rockville, MD, p. 337-344, 1996, Nov.4-5
- [43] D. Dakopoulos, N. Bourbakis, “A 2D vibration array as an assistive device for visually impaired,” *IEEE Int. Conf. on BIBE07*, Boston, MA, vol. II, pp. 930-937, 2007, Oct. 15-17

- [44] N. Bourbakis (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.
- [45] N. Bourbakis, P. Kakumanu, "Skin-based face detection- extraction and recognition of facial expressions", *Applied Pattern Recognition*, book chapter, Vol.91, 2008
- [46] D. Dakopoulos, N. Bourbakis, "Preserving visual information in low resolution images during navigation of blind", 1st International Conference on Pervasive Technologies Related to Assistive Environments, Athens, Greece, 2008, July 15-19
- [47] P. Mengel, G. Doemens, L. Listl, "Fast range imaging by CMOS sensor array through multiple double short time integration (MDSI)", *Proc. IEEE International Conference in Image Processing (ICIP 2001)*, Thessaloniki (Greece), Oct.7-10, 2001, pp.169-172
- [48] P. Mengel, L. Listl, B. Konig, C. Toepfer, M. Pellikofer, W. Brockherde, B. Hosticka, O. Elkhali, O. Schrey, W. Ulfing, "Three-Dimensional CMOS image sensor for pedestrian protection and collision mitigation", J. Valldorf and W. Gessner, Eds., *Advanced microsystems for automotive applications 2006*. Berlin: Springer, 2006. (VDI-Buch), pp. 23-39
- [49] O. Elkhali, O.M. Scherey, P. Mengel, M. Petermann, W. Brockherde, "A 4x64 pixel CMOS image sensor for 3-D measurement applications", *IEEE Journal of Solid Circuits*, vol. 39, 7, pp. 1208-1212, 2004

7. DISCUSIÓN GENERAL DE RESULTADOS

La ceguera es una gran deficiencia que afecta en todo el mundo a 39 millones de personas. Se calcula que cada cinco segundos una persona se queda ciega y que cada minuto un niño pierde la vista. Eso implica que cada año más de siete millones de personas se ven afectadas por esta discapacidad, según la OMS [1]. El mayor aumento de en los países desarrollados se debe a enfermedades relacionadas con la edad, como la degeneración macular asociada a la edad (DMAE), la catarata (corregible mediante intervención quirúrgica), el glaucoma y la retinopatía diabética.

En un informe de 2002 de Naciones Unidas [2] se indica que en 2050 España será el país más envejecido del mundo, ya que más del 44% de la población española será mayor de 60 años, por lo que España será el país donde más crecerá porcentualmente la ceguera asociada a la edad.

La razón fundamental y la especial motivación que ha llevado al desarrollo de esta tesis es que las personas invidentes puedan disponer de un dispositivo que les ayude a desenvolverse con más autonomía, un sistema que les permita detectar obstáculos, algo que no pueden hacer al cien por cien con las herramientas de que disponen actualmente.

Como se ha avanzado en la introducción de esta tesis, desde los años 50 se ha desarrollado más de 45 dispositivos de navegación basados en diferentes tecnologías, como los ultrasonidos, la visión artificial, los mapas electrónicos, los sistemas de estereovisión, etc... Pero todos ellos no han pasado más que de la versión de prototipo. Los problemas que han tenido estos dispositivos han sido diversos. El principal de ellos es que los sistemas de representación de la información recibida del entorno son muy confusos y requieren que el usuario tenga una alta concentración para descifrar dicha información, lo que le impide detectar cualquier otro tipo de mensaje o estímulo que pueda venir del entorno real.

Un ejemplo son los sistemas basados en representación del entorno mediante voz sintética, que consisten en que el invidente es guiado mediante voz sintetizada por un ordenador. En este caso su principal problema está en que el procesado de la información es lento y requiere mucha atención para que el usuario interprete el entorno real estando en movimiento, ya que debería

moverse muy lentamente para que el sistema procesara la información y que ésta le fuera útil.

También hay otros tipos de dispositivos basados en sensación táctil. Algunos sistemas representan el entorno mediante una matriz de puntos, que se detecta mediante la palma de la mano o mediante sensores que hay que introducirse en la boca y genera mediante impulsos el mapa del entorno real. Este tipo de dispositivos sensoriales requieren también de un alto grado de concentración, lo que dificulta al usuario poder captar información adicional del entorno.

Otra de las desventajas de la mayoría de estos dispositivos es que se encuentran en fase de desarrollo o no han superado más allá de la etapa de prototipo, por lo que tienen un coste muy elevado.

Ello nos lleva a concluir que, si comparamos el sistema analizado en la presente tesis con los dispositivos que hemos detallado anteriormente, éste es muy fiable y preciso. De hecho, el índice de error del láser CMOS 1x64 es inferior al 1% para la detección de los obstáculos en el campo cercano de hasta 5 metros y tiene una respuesta que se aproxima al tiempo real.

Otra de las principales ventajas que ofrece es que la representación del entorno real se realiza mediante un sonido corto, bien tolerado por el usuario, una especie de ‘clic’. Está demostrado que el cerebro humano procesa con mayor rapidez ese tipo de sonidos que no necesitan tanta atención como las sensaciones táctiles.

Se ha comprobado que uno de los aspectos que genera mayor aceptación entre los usuarios es que desde que se captura el entorno hasta que el invidente escucha el sonido virtual posicionado en el lugar donde se encuentra el obstáculo no llega a trascurrir un segundo, lo que se puede considerar prácticamente como tiempo real. Por ello, el usuario del sistema tiene total libertad de movimiento, sin tiempo de espera en el procesado de la información que recibe del entorno.

El sistema ha demostrado ser estable y fiable tanto en espacios abiertos como en interiores, con gran aceptación y adaptabilidad de las personas invidentes. Una vez realizadas las pruebas, se ha detectado que las personas no invidentes de nacimiento tenían mayor facilidad de adaptación al sistema que las personas invidentes de nacimiento. Se comprobó que con entrenamiento, todos los usuarios mejoraban en la localización de obstáculos. Una vez

finalizadas las pruebas, todos ellos lo consideraron una herramienta que mejoraría considerablemente su adaptación al entorno real.

7.1 Referencias

- [1] OMS. Organización Mundial de la Salud. Organismo Internacional para la Prevención de la Ceguera (IAPB).
Disponible en: <http://www.who.int/mediacentre/news/releases/pr79/es/>
- [2] United Nations. Population Division. World Population Ageing 1950-2050. 2002. Disponible en: [http:// www.who.int/](http://www.who.int/).

8. CONCLUSIONES

Podemos concluir en la presente tesis que se han podido alcanzar los objetivos planteados en la fase de investigación.

1. El dispositivo utilizado como herramienta de captación del entorno real basado en una serie de sensores láser CMOS 1x64, ha demostrado ser muy preciso y tolerante a fallos. Se han detectado errores inferiores al 1% en las medidas realizadas, por lo que se considera un dispositivo apto para la captación del campo cercano de entre 0,5 y 5 metros de distancia. La respuesta del sistema con respecto al procesado y salida de información podemos considerarla como tiempo real, ya que desde la captación del entorno hasta la salida del dispositivo transcurre un intervalo de tiempo inferior a 1 segundo.
2. Las pruebas realizadas para la localización de sonidos virtuales en el espacio demostraron que la localización de un sonido en distancia y azimut no depende del tipo de material simulado en el sonido emitido. También se pudo concluir que el efecto de la reverberación no aumenta la precisión en la localización del sonido.
3. La utilización de Head Related Transfer Functions (HRTF) para el posicionamiento virtual de los objetos simulando el entorno real se ha demostrado que es viable, ya que se ha obtenido muy buenos resultados en las pruebas con personas invidentes. De hecho, se ha comprobado que la mejora y la precisión en la detección de obstáculos se conseguía mediante el entrenamiento. También se llegó a la conclusión de que la curva de aprendizaje era menor en personas que no siendo invidentes de nacimiento habían perdido la visión con posterioridad debido a una enfermedad o accidente.
4. Finalmente el dispositivo ha demostrado que es capaz de localizar los obstáculos con gran precisión de forma fiable, para que una persona

invidente sea capaz de detectar dicho obstáculo a una distancia de entre 0,5 y 5 metros. Eso supone superar considerablemente el radio de alcance de un bastón, que no es más que un metro o metro y medio como máximo. También ofrece una ventaja con respecto a los obstáculos que se encuentran a la altura de la cabeza y tronco de una persona invidente, como puede ser un toldo de una tienda o un camión cruzado en una acera. Según constatan los propios invidentes, los bastones pasan sin dificultad por debajo de ese tipo de obstáculos sin detectarlos, lo que en muchos casos puede provocarles una colisión.

Como comentario final, debemos destacar que todas personas invidentes que realizaron las pruebas mostraban su interés de disponer del dispositivo para seguir entrenando y utilizarlo en su vida diaria. Un hecho que, obviamente, supone una gran motivación para seguir investigado en esta línea.

Como futuros trabajos de investigación en la mejora del sistema se plantean las siguientes líneas:

1. Utilizar más dispositivos 1x64 3D CMOS para poder adquirir varios planos del entorno real haciendo un barrido en vertical además del horizontal. Ello permitiría detectar sonidos virtuales también en altura.
2. Integrar un sistema de reconocimiento de textos para que cuando la persona invidente estuviera delante de un rótulo, documento o teniendo en su mano la carta de menú de un restaurante, el dispositivo leyera esa información.
3. Integrar un sistema de reconocimiento facial, que sea capaz de detectar y reconocer a las personas que hay en el entorno cercano del usuario invidente y que incluso sea capaz de posicionarlo con su nombre en el espacio.

