ATAD: Assistive Technology for an Autonomous Displacement

TESIS DOCTORAL

Autor:

Pablo Revuelta Sanz

Directores:

Belén Ruiz Mezcua / José M. Sánchez Pena

DEPARTAMENTO

Tecnología Electrónica

Leganés, Junio 2013

Universidad Carlos III de Madrid

TESIS DOCTORAL

ATAD: Assistive Technology for an Autonomous Displacement

Ayuda Técnica para la Autonomía en el Desplazamiento

Autor: **Pablo Revuelta Sanz**

Directores: **Belén Ruiz Mezcua / José M. Sánchez Pena**

Firma del Tribunal Calificador:

		Firma
Presidente:	(Nombre y apellidos)	
Secretario:	(Nombre y apellidos)	

Calificación:

Leganés, de de

Agradecimientos Personales

Un trabajo así no se hace sólo, ni tampoco lo hace una sola persona. Por ello, he de agradecer a las siguientes personas el apoyo prestado, la colaboración y los consejos que han hecho de una idea loca una tesis doctoral.

Aunque sean muchas más las personas que de un modo u otro han intervenido en mi vida durante estos años, en concreto hay personas que han dedicado una parte de su tiempo a esta tesis, y quiero mostrarles mi agradecimiento:

- Belén Ruiz Mezcua
- José M. Sánchez Pena
- Jean-Phillippe Thiran
- Bruce N. Walker
- Luigi Bagnato
- Myounghoon “Philart” Jeon
- Anus
- Sil
- Tete
- Mi madre
- Mi hermana
- Mi padre
- Javier Jiménez Dorado
- Mª Jesús Varela
- Ignasi Terraza
- Mayca Cruz
- Alfredo Sanz Hervás
- Virginia Yagüe
- Soledad Mochales
- Luís Hernández
- Cristina R.-Porrero
- Enrique Varela
- Rosa Lucerga
- Carlos Guindos Jiménez
- Ana Patricia Espejo
- Carmen Bonet Borras
- Fco. Javier Ruiz Morales
- Pachi García Lasheras
- Enrique Sainz de Murieta y Rodeyro
- Carmen María Massanet Arco
- Ashley Henry
- Prince Gibbons
- Riley Winton
- Jonathan Schuett
- Darell Shackelford
- Linda Sue Jonhson
- Vincent Martin
- Reginald Reece
- Annie Kim
- Hannah Fletcher
- Cynthia Halloway
- Dereck Price
- Thom Gable
- Hannah Cauley
- Theresa Watkins
- Charmia Dixon
- Rajnikant Mehta
- John Smith
- Tomica Savoy
- Melissa Imtiaz
- Hyoun Sally Park
- Hyeonjin
- Gedeion
- Tommy Thomas
- Diana Anglin
- James Baker
- Arrion
- Sang Hun
- Gerard
- Nathen Steele
- Ze Chen

Agradecimientos Institucionales

Comunidad de Madrid, por la financiación no. S2009/ESP-1781).

Universidad Carlos II de Madrid, por la Beca cofinanciada 2009-2013

Centro Español de Subtitulado y Audiodescripción, por la Beca cofinanciada 2009-2013

École Polytechnique Fédérale de Lausanne, por Estancia e infraestructuras

School of Psychology, Georgia Institute of Technology, por Estancia e infraestructuras.

Center for the Visually Impaired, GA, por sus contactos

APTENT, por el subtitulado del video final

Table of Contents

Annex I: MIDI Messages 311

Annex II. User Tests over Virtual Reality 321

Annex III. Experts Training and Testing 333

Preliminary Terminology Note

The term “technical aid” has been widely used for pointing to technology applied to improve the life quality of people.

This term has been changed to that of *Assistive Product* (AP from now on) in the regulation ISO 999:2007.

In this text, we will use this last term to design such application of technology and, hence, our proposal.

Resumen de la Tesis

El proyecto desarrollado en la presente tesis doctoral consiste en el diseño, implementación y evaluación de una nueva ayuda técnica orientada a facilitar la movilidad de personas con discapacidad visual.

El sistema propuesto consiste en un procesador de estereovisión y un sintetizador de sonidos, mediante los cuales, las usuarias y los usuarios pueden escuchar un código de sonidos mediante transmisión ósea que les informa, previo entrenamiento, de la posición y distancia de los distintos obstáculos que pueda haber en su camino, evitando accidentes.

En dicho proyecto, se han realizado encuestas a expertos en el campo de la rehabilitación, la ceguera y en las técnicas y tecnologías de procesado de imagen y sonido, mediante las cuales se definieron unos requisitos de usuario que sirvieron como guía de propuesta y diseño.

La tesis está compuesta de tres grandes bloques autocontenidos: (i) procesado de imagen, donde se proponen 4 algoritmos de procesado de visión estéreo, (ii) sonificación, en el cual se detalla la propuesta de transformación a sonido de la información visual, y (iii) un último capítulo central sobre integración de todo lo anterior en dos versiones evaluadas secuencialmente, una software y otra hardware.

Ambas versiones han sido evaluadas con usuarios tanto videntes como invidentes, obteniendo resultados cualitativos y cuantitativos que permiten definir mejoras futuras sobre el proyecto finalmente implementado.

Abstract

The project developed in this thesis involves the design, implementation and evaluation of a new technical assistance aiming to ease the mobility of people with visual impairments. By using processing and sounds synthesis, the users can hear the sonification protocol (through bone conduction) informing them, after training, about the position and distance of the various obstacles that may be on their way, avoiding eventual accidents.
In this project, surveys were conducted with experts in the field of rehabilitation, blindness and techniques of image processing and sound, which defined the user requirements that served as guideline for the design.

The thesis consists of three self-contained blocks: (i) image processing, where 4 processing algorithms are proposed for stereo vision, (ii) sonification, which details the proposed sound transformation of visual information, and (iii) a final central chapter on integrating the above and sequentially evaluated in two versions or implementation modes (software and hardware).

Both versions have been tested with both sighted and blind participants, obtaining qualitative and quantitative results, which define future improvements to the project.

1. Introduction

The work carried out in this Ph.D. doctoral Thesis involves the design, implementation and validation of an Assistive Product (AP) to help blind people in their mobility, to circumvent obstacles in known or unknown environments.

We will get into the AP world, the image processing, the psychoacoustics, electronics and user tests. It is, then, a multidisciplinary project with some objective and radical divisions, which will be respected in the memory structure.

Thus, in this first chapter, we will only focus on the general statement of the problem we want to address, the AP for blind people. The rest of chapters will present the corresponding state of the art and bibliography, for an easier comprehension.

1.1. Introduction to Assistive Products

An "Assistive Product" (AP) is "*any product (including devices, equipments, instruments, technology and software) produced specially or available in the market, to prevent, compensate, control, reduce or neutralize deficiencies, limitations in activity and restrictions in participation*" [1]. This means a specific application of technology to enhance the quality of living of people. In such a wide definition, almost every technology could be taken as AP, like cars, computers or houses.

In practical terms, the concept of AP is applied to technology applied to solve problems in people with some kind of temporal or permanent disability.

The WHO states that an AP's is "any product (including devices, equipment, instruments and software), especially produced or generally available, used by or for persons with disability:

- for participation;
- to protect, support, train, measure or substitute for body functions / structures and activities; or
- to prevent impairments, activity limitations or participation restrictions" [2].

Thus, the following items are excluded [2]:

- Items used for the installation of assistive products;
- Solutions obtained by combinations of assistive products which are individually classified in the classification;
- Medicines;
- Assistive products and instruments used exclusively by healthcare professionals;
- Non-technical solutions and services, such as personal assistance, guide dogs or lip-reading;
- Implanted devices;

We can track the AP even from the beginning of History [3]. However, it is a modern concept, which was firstly established by the ISO/TC173/SC2 in 1983 [2].

The ISO 9999 norm establishes, in its last issue [1], the following classification, at “class” level:

04	AP for personalized medical treatment
05	AP for training/learning of skills
06	Prosthesis
09	AP for care and personal protection
12	AP for personal mobility
15	AP for domestic activities
18	Adaptations for homes
22	AP for communication and information
24	AP for objects and devices manipulation
27	AP to improve environment, tool and machines
30	AP for entertainment

Table 1.1. ISO 9999 classification for AP. Class level.

In general terms, AP’s are the more and the more used by this collective and we can easily see blind people with white sticks or Braille computers, deaf people with cochlear implants or audiphones, physically impaired people with wheelchairs, etc.

From an economical point of view, the limited market of application of AP’s has worked as a constraint for their development under a capitalist and free market economic system: the limited population of disabled people is an impediment to develop scale economics in production [4]. This limitation affects, consequently, to design. Hence, both prototypes and commercial products use to be very expensive. This is the reason of the usual failure that most of the novel AP’s have to confront when coming in the market. Because of that, AP’s are usually designed and built-up with commercial technology. Thus this field almost never works as an avant-garde in technology development, as banking, military or luxury technology use to act. Results of these constraints are shown, for example, in the article of R. M. Mahoney, where he analyzes costs and number of sold products of different robotic products for rehabilitation [5].

We can clearly see the importance of “extra-market” motivations to develop and distribute this kind of technological products, as proposed in [6]. Hence, social institutions should invest in this and many other social-valued fields of research, development and innovation. These collectives cannot wait to the market, which may not even arrive.

1.1.1. Specific Application Field

By definition, there is no a global AP, that is, an AP applicable to any kind of problem. Technology is applied, since first times, to solve problems, but specific problems. The design is constrained by the final application and that one must be taken into account if the AP wants to be applicable. Then, we have to focus on a specific problem and, hence, collective, to be able to design and implement a practical solution. Maybe we could accept the personal computer as the only generalized AP that it has ever been built.

Anyhow, here we can see the capital importance of the choice of the target population and problem. This is an arbitrary choice, since there are many and subjectively evaluated problems in the life of disabled people.

1.2. Target Population

As said before, the choice of the target population will determine the design, implementations, user tests and, of course, practicality of the new device.

1.2.1. Disability Models

When we want to deal with disability, we are assuming a disability model, that is, a conceptual approach to disability, which generates specific roles for both disabled and non-disabled people.

These models have been summarized by Ladner [7]:

- Medical Model: disabled people are patients who need a treatment or cure.
- Rehabilitation Model: disabled people are clients. They need AP's at everyday life. Technology is usually paid by the users.
- Special Education Model: focused on children with disabilities, they use to need special cares and aids in their education.
- Legal Model: disabled people are citizens with the rights and responsibilities, which are ensured by laws. AP's are given for free.
- Social Model: disabled people are part of the diversity of life, not needing special care, but special technology to partake in social life. AP's are partially or totally paid by individuals or it's free, if some law prescribes it.

The proposed models are, as it can be easily seen, overlapped. Hence, we can choose some of them, if needed, to afford the project. Of course, some of them, like the medical of the legal approaches, are out of the ambit of this study. As proposed, this AP is not specifically focused on special education. In other words, even if it is important to take children into account to evaluate or test any AP, this AP is not designed, as it will be shown, for educational purposes.

Thus, we will take the rehabilitation and the social model as guides to approach the visual disability ambit.

1.2.2. Looking at the Blindness

We have focused the work on the visual impaired people, and more specifically, on the mobility problems and challenges this collective must confront every time they go out of known environments.

“Blindness” is a very common word to design the situation where people cannot see. But the medical definition must be more accurate than that.

The first problem we find when searching for a more accurate definition is the threshold between “blindness” and “low vision”. A report from the World Health Organization (WHO) proposes the following definition for “low vision” [8]:

> A person with low vision is one who has impairment of visual functioning even after treatment and/or standard refractive correction, and has a visual acuity of less than 6/18 to light perception, or a visual field of less than 10 degree from the point of fixation, but who uses, or is potentially able to use, vision for planning and/or execution of a task.

In the same report, it is stated that "the current definition does not make a distinction between those who have "irreversible" blindness (no perception of light) and those that have light perception but are still less than 3/60 in the better eye." [8].

Finally, we can find a brief description of this characteristic [9]:

> Blindness is the inability to see. The leading causes of chronic blindness include cataract, glaucoma, age-related macular degeneration, corneal opacities, diabetic retinopathy, trachoma, and eye conditions in children (e.g. caused by vitamin A deficiency). Age-related blindness is increasing throughout the world, as is blindness due to uncontrolled diabetes. On the other hand, blindness caused by infection is decreasing, as a result of public health action. Three-quarters of all blindness can be prevented or treated.

More in detail, we present in table 1 the last classification given for blindness, in terms of light perception [8].

Category	Worse than:	Equal or better than:
Mild or no visual Impairment 0		6/18 3/10 (0.3) 20/70
Moderate visual impairment 1	6/18 3/10 (0.3) 20/70	6/60 1/10 (0.1) 20/200
Severe visual impairment 2	6/60 1/10 (0.1) 20/200	3/60 1/20 (0.05) 20/400
Blindness 3	3/60* 1/20 (0.05) 20/400	1/60* 1/50 (0.02) 5/300 (20/1200)
Blindness 4	1/60* 1/50 (0.02) 5/300 (20/1200)	Light perception
Blindness 5	No light perception	
9	Undetermined or unspecified	

*Or counts finger (CF) at 1 meter.

Table 1.2. Blindness classification by levels.

Thus, "blindness" is specifically every state that fits in categories 3, 4 or 5.

These categories are useful to design a good enough approach to this collective, which has, as seen, many different capabilities and it is obviously not homogeneous. This condition will be discussed in Chapter 2.

1.2.3. Facts about blindness

Regarding the last sheet about blindness from the WHO [10], we take the following overall sightsee:

- About 314 million people are visually impaired worldwide; 45 million of them are blind.

- Most people with visual impairment are older, and females are more at risk at every age, in every part of the world.
- About 87% of the world's visually impaired live in developing countries.
- The number of people blinded by infectious diseases has been greatly reduced, but age-related impairment is increasing.
- Cataract remains the leading cause of blindness globally, except in the most developed countries.
- Correction of refractive errors could give normal vision to more than 12 million children (ages five to 15).
- About 85% of all visual impairment is avoidable globally.

We can find in these data the magnitude of that problem, which affects to almost 3% of the population (regarding estimations from the WHO).

An important aspect to point out is that visually impairment generates difficulties for moving or accessing information, but this can generate a disability if the environment doesn’t support the people with these difficulties. Thus, the visual disability is a combination of some physiological problems and the technological, legal, economical and cultural environment.

1.2.3.1. Distribution

Since blindness has not only biological but also socio-economical aspects (and then, it is not a random variable among the human beings), it is interesting to perform a more detailed analysis about the distribution all over the world.

Figure 1 presents the prevalence of blindness (regarding the WHO classification) in the world [11].

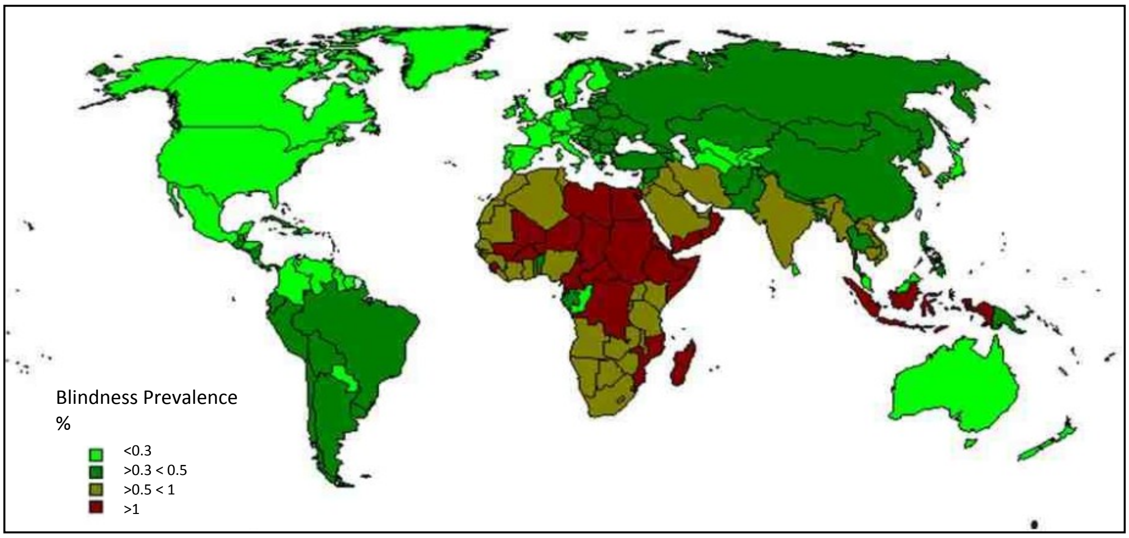


Fig. 1. 1. Blindness prevalence in the world.

More in detail [12]:

	Prevalence of blindness (%)			No. of blind persons (millions)		
Subregion	<15 years of age	15–49 years	>=50 years	<15 years of age	15–49 years	>=50 years
Afr-D	0.124	0.2	9	0.191	0.332	3.124
Afr-E	0.124	0.2	9	0.196	0.336	3.110
Amr-A	0.03	0.1	0.4	0.021	0.114	0.560
Amr-B	0.062	0.15	1.3	0.085	0.369	0.937
Amr-D	0.062	0.2	2.6	0.017	0.075	0.241
Emr-B	0.08	0.15	5.6	0.039	0.117	0.920
Emr-D	0.08	0.2	7	0.043	0.146	1.217
Eur-A	0.03	0.1	0.5	0.021	0.204	0.713
Eur-B1	0.051	0.15	1.2	0.020	0.136	0.462
Eur-B2	0.051	0.15	1.3	0.009	0.043	0.090
Eur-C	0.051	0.15	1.2	0.021	0.192	0.822
Sear-B	0.083	0.15	6.3	0.102	0.332	3.779
Sear-D	0.08	0.2	3.4	0.390	1.423	6.530
Wpr-A	0.03	0.1	0.6	0.007	0.070	0.315
Wpr-B1	0.05	0.15	2.3	0.162	1.166	6.404
Wpr-B2	0.083	0.15	5.6	0.041	0.120	1.069
Wpr-B3	0.083	0.15	2.2	0.002	0.006	0.017
World				**1.368**	**5.181**	**30.308**

Table 1.3. Prevalence of blindness by regions.

Where the different subregions and countries related to them are shown in the next table [12]:

Subregion	Studies
Afr-D	Surveys from 13 countries (Benin, Cameroon, Cape Verde, Equatorial Guinea, Gambia, Ghana, Mali, Mauritania, Niger, Nigeria, Sierra Leone, Sudan, Togo)
Afr-E	Surveys from 6 countries (Central African Republic, Congo, Ethiopia, Kenya, South Africa, United Republic of Tanzania)
Amr-A	Surveys from 1 country (United States of America)
Amr-B	Surveys from 3 countries (Barbados, Brazil, Paraguay)
Amr-D	Survey from 1 country (Peru)
Emr-B	Surveys from 4 countries (Lebanon, Oman, Saudi Arabia, Tunisia)
Emr-D	Survey from 1 country (Morocco)
Eur-A	Surveys from 7 countries (Denmark, Finland, Iceland, Ireland, Italy, Netherlands, United Kingdom)
Eur-B1	Surveys from 2 countries (Bulgaria, Turkey)
Eur-B2	Survey from 1 country (Turkmenistan)
Eur-C	No population-based surveys were identified
Sear-B	Surveys from 4 countries (Indonesia, Malaysia, Philippines, Thailand)
Sear-D	Surveys from 4 countries (Bangladesh, India, Nepal, Pakistan)
Wpr-A	Surveys from 1 country (Australia)
Wpr-B1	Surveys from 2 countries (China and Mongolia)
Wpr-B2	Surveys from 3 countries (Cambodia, Myanmar, Viet Nam)
Wpr-B3	Surveys from 2 countries (Tonga and Vanuatu)

Table 1.4. Relation between subregions and countries.

It is not the scope of this work to perform a detailed and sociological analysis of this distribution, but it is interesting to notice that blindness is much more prevalent in impoverished countries. This fact will limit the applicability of whichever AP is designed and, therefore, ours too.

1.2.3.2. Causes of blindness

To understand the group to which we are addressing the present work, we have to analyze the causes of blindness [12].

Region	Cataract	Glaucoma	AMD*	Corneal opacities	Diabetic retino-pathy	Childhood blindness	Trachoma	Oncho-cerciasis	Others
Afr-D	50	15		8		5,2	6,2	6	9,6
Afr-E	55	15		12		5,5	7,4	2	3,2
Amr-A	5	18	50	3	17	3,1			3,9
Amr-B	40	15	5	5	7	6,4	0,8		20,8
Amr-D	58,5	8	4	3	7	5,3	0,5		13,7
Emr-B	49	10	3	5,5	3	4,1	3,2		22,2
Emr-D	49	11	2	5	3	3,2	5,5		21,3
Eur-A	5	18	50	3	17	2,4			4,6
Eur-B1	28,5	15	15	8	15	3,5			15
Eur-B2	35,5	16	15	5	15	6,9			6,6
Eur-C	24	20	15	5	15	2,4			18,6
Sear-B	58	14	3	5	3	2,6			14,4
Sear-D	51	9	5	3	3	4,8	1,7		22,5
Wpr-A	5	18	50	3	17	1,9	0,025		5
Wpr-B1	48,5	11	15	3	7	2,3	6,4		6,8
Wpr-B2	65	6	5	7	3	3,6	3,5		6,9
Wpr-B3	65	6	3	3	5	9,5	4,3		4,2
World	**47,8**	**12,3**	**8,7**	**5,1**	**4,8**	**3,9**	**3,6**	**0,8**	**13**

*AMD, age-related macular degeneration.

Table 1.5. Causes of blindness by regions.

Some relevant information can be extracted from this table. In Europe, the most important causes of blindness (AMD, Glaucoma and diabetic retinopathy) are strongly related to age. This effect can also be seen in table 1.5, where the prevalence in lower ages is around 0.1%, but when we regard to that over 50 years old, it raises till 0.5%. The same result appears when regarding the U.S.A. (since socio-economic conditions are, mostly, the same).

This means that most of the people, in Europe and North America, who are blind, have seen more or less in some period of their lives.

There is another preliminary conclusion: the AP's for blind people have to take into account elderly people, since they are the most important group in this collective.

Although most of the causes of blindness have a chirurgical or medical treatment (cataract still causes the 47.9% of blindness cases in impoverished countries [13]), we can find a group of blind people, also in the rich countries who could get advantage of AP's to improve their lives.

1.3. Orientation and Mobility

Moving in a 3D space requires orientation and navigation capabilities [14]. For that purpose, we are able to gather, interpret and build up knowledge about our environment in a multifaceted skill [15].

The collective of skills, techniques and strategies used by blind people to travel independent and safely are known as Orientation and Mobility (O&M) [16-19].

As stated at the beginning of this chapter, our proposal is focused on mobility, thus, class 12 in table 1.6. More in detail, this class presents the next sub-classification [1]:

12 03	AP to walk managed by an arm
12 06	AP to walk managed by both arms
12 07	Accessories for walking AP
12 10	Cars
12 12	Adaptations to cars
12 16	Motorbikes
12 18	Cycles
12 22	Wheelchair with manual propulsion
12 23	Wheelchair with motor propulsion
12 24	Accessories for wheelchairs
12 27	Vehicles
12 31	AP for transference and turn
12 36	AP for elevation
12 39	AP for orientation

Table 1.6. Classification of AP for mobility. Sub-class level.

We can get into the final level of classification of this norm, specifically for the Orientation tools [1]:

12 39 03	White/Tactile canes
12 39 06	Electronic AP for orientation
12 39 09	Acoustic AP for navigation (sonic beacons)
12 39 12	Compass
12 39 15	Relief maps
12 39 18	Tactile orientation materials

Table 1.7. AP for orientation classification. Lower level.

It is crucial to remark the difference between orientation and mobility, although they use to appear even in the same acronym as O&M. Regarding the specialists C. Martínez and K. Moss, Orientation means to “know where he is in space and where he wants to go”, while mobility is to be “able to carry out a plan to get there” [20]. A similar classification was proposed by Petrie *et al.* [21] as “macronavigation” and “micronavigation” respectively. The first definition points to a general idea of the situation in a geographical sense. The second one takes into account how to move, to travel, given a position and a direction. We will focus in this works on mobility tools, usually referred to as Electronic Travel Aids (ETA’s).

Blind people can use the extra information provided by these devices, gathered with other environmental information, to supplement the visual limited information, achieving a successful navigation [22]. We will review the proposed devices for this purpose in chapter 2.

1.4. Motivation

We find two cores of motivation to propose a novel assistive product:

- Blind people (as well as other disabled collectives) suffer a lack of mobility right, in practical terms. Even if architecture and urbanism has implemented some solutions, many unexpected obstacles and changes in urban furniture act as a barrier to a practical enjoyment of this right. This has been time ago identified as one of the most important functional limitations of blind and visually impaired people [23, 24]. This is an ethical problem more than a technical one, and its fundaments are discussed in detail in [4], from a liberal point of view.
- We found problems in the analyzed AP's, regarding usability, complexity, weight and, above all, price. This field must keep proposing solutions to the mobility of blind people.
- We want to provide to the visually impaired people an AP which may help them gaining independence in their daily life displacements. For that, a design of new and light image processing algorithms will be given, as well as new sonification proposals, in order to gather all these algorithms and make them running in usable and cheap hardware.

More in detail, we can find the following limitations, regarding [25], we find limitations in proposed AP's regarding:

- the types, amounts, and accuracy of information they can provide,
- the types of environments in which they can function, and
- their user interface structure/operating procedures.

As it has been seen in the 1.2.2 section, blindness is an enough important aspect of our social reality.

Specifically talking about systems for mobility and even "vision", we found that those systems are used to be the most expensive ones because the high technology and knowledge involved in their design and fabrication. Moreover, these systems use to be heavy and awkward, using important batteries or cameras. It is true that some of them were proposed several years ago, when microelectronic was not as small and cheap as it is nowadays. However, even in some novel proposals, we find the aforementioned limitations in terms of price and usability.

There are challenges found focused on some important aspects of technology applied to impairments [26]:

- Most of technology is addressed to above average intelligence.
- There is an unmet need of low-cost solutions.
- Many products stay in a prototype state, without achieving the users' world.

1.4.1. The Assistive Product Interaction Model

Assistive products interact with users in a complex way. If we want to propose a novel AP with a high-added value, we must to define a paradigm of interaction between such AP and the potential user. This is the so called Interaction Model (IM).

The first IM proposed is found in 1991 [27, 28]. The proposed model was called the Human-Environment/Technology Model (HETI), defining a system composed of human, environment and technology. The HP is done, obviously, with all the cognitive capabilities of the user. Figure 1.2 presents a scheme of such model, following [27].

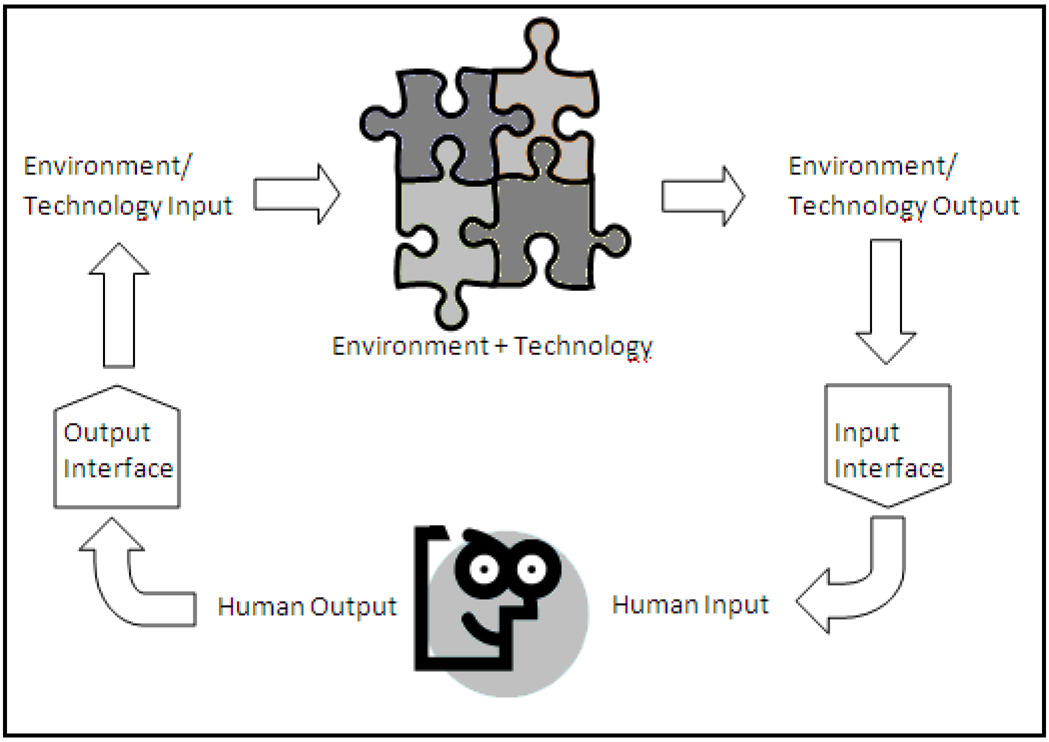


Fig. 1.2. HETI model.

This model has been completed in [29], introducing activity and tasks factors, since humans don't interact in the same way with technology depending on the specific activity. In this version of the model, the context is composed by human, activity and technology.

We will focus in this approximation our proposal, since the project is specifically addressed to a very particular task, i.e., moving in unknown environments.

The proposed interfaces will be discussed later. For instance, we are interested in the human cognitive capabilities, to achieve our goal, which is the issue discussed in the following section.

1.4.2. The possibilities of Mental Imagery

As stated in [16], "spatial concepts are used to form a conceptual understanding, or cognitive map, of a given environment".

The proposal presented in this research aims to “induce” a mental image by means of non-standard paths, since those paths are, somehow, interrupted in blind people. We have to deal with the so-called mental imagery to help to understand the environment.

This process has been widely studied. Initial studies reporting how the brain works when dealing with mental images start in the earliest 80’ [30-32], while in psychological terms, studies can be found much earlier [33]. These early works propose that the mental images are created not only from the eyes, but from other sources. In [30], it is reported that mental images can represent spatial relationships that were not explicitly in the forming image. We can use, like when listening to an orientation scheme, a spatial reasoning to form such mental images [30]. Moreover, mental images respond to a specific viewpoint [30]. Hence, we can assume that a subjective point of view is intrinsic to mental images formation.

In the following years, we start to find the first evidences of hemispheric asymmetries and brain specialization in mental imagery [31, 32].

Psychologists have shown how the brain works to build up mental images with some more material than that coming from the eyes [34, 35].

The following figure is a simple experiment to see the effect of the blind spot. Shutting the right eye (the experiment is symmetric for both eyes), you look at the star at 30-50 cm and you will stop seeing the circle at some distance. The brain, however, fills up this lack of information if you repeat the experience with both eyes opened.

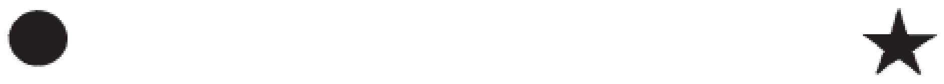


Fig. 1.3. Blind spot experiment [34].

In the last decades, neuroimaging techniques like Magnetic Resonance Imaging (MRI) or functional MRI (fMRI) carried to more accurate 2D and 3D images of the brain, even working.

With those techniques, a new empirically verifiable hypothesis rises up: “Visual imagery is the ability to generate percept-like images in the absence of retinal input” [36]. Moreover, “visual imagery can be dissociated from visual perception” [37]. This is the effect of the so-called “cross-modal plasticity” of the brain [38]. This plasticity was already suggested one decade before, as in [39] for example. We can find, in this study some of the first evidences of such effect. Figure 1.4 shows two brain maps extracted but positron emission tomography (PET) of a brain constructing mentally a 3D object following some verbal instructions and the same brain reposing (performing no task).

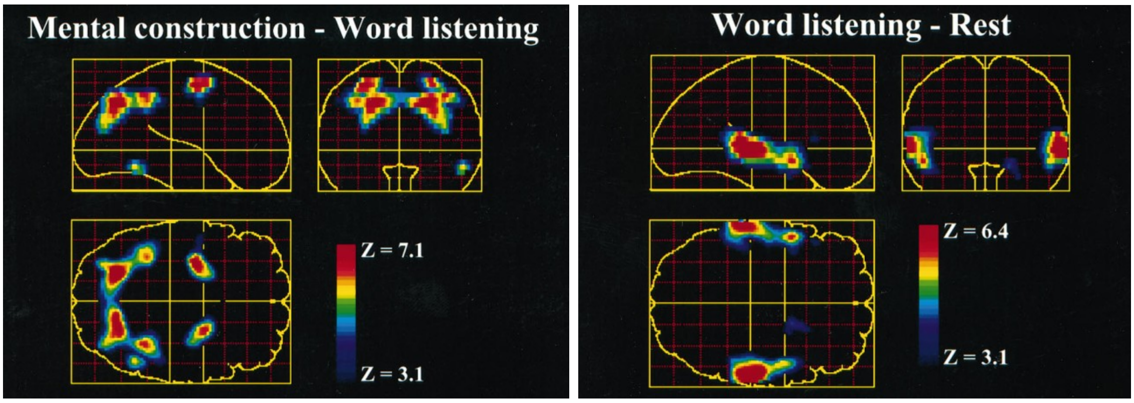


Fig.1.4. Brain spatial imagery activation by means of verbal instructions [39].

Not only verbal instructions activate some visual areas. In [40], it is shown how sounds of objects can activate the same visual areas in blind and sighted persons (however, with a different intensity).

In the late 90's, some researchers started to analyze specifically the brain of blind people. In the study presented by Büchel *et al.* [41], finding relevant differences between early and late blindness. These differences show that vision in the earlier steps of development help forming structures in the visual cortex [41]. These data were obtained when reading in Braille. In the same study, they argue that non-primary areas are more suitable to reorganize their functionality and connections that primary areas.

However, this discussion was not closed yet. The RMI and fRMI start to provide very accurate information about these effects observed by PET and other older technology. For example, in [42], researchers broach the conclusions of [41] finding that primary visual areas (PVAs) are activated, even in congenital blind people, in absence of stimuli:

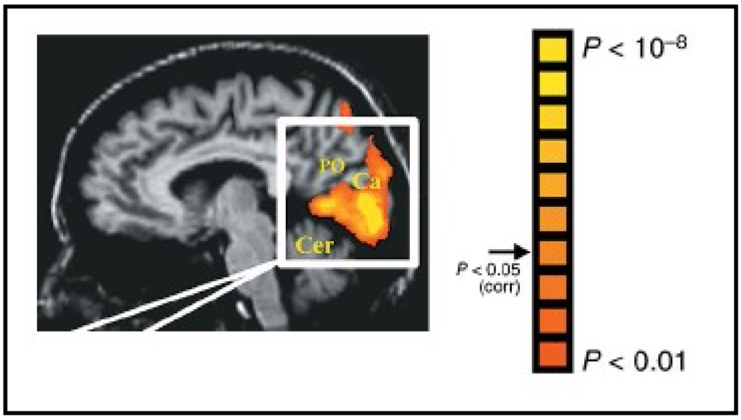


Fig. 1.5. Detail of PVAs activation, during a verbal memory task [42].

Comparative studies between blind and sighted people have been widely reported. In [43], some similarities and differences in the response of several sensorial areas are found. As general conclusion of this study regarding our goal, we find a more developed cross-modal adaptation of visual areas in blind people regarding the sighted group. This effect can be found, as stated in [44, 45], even in sighted people with temporal visual deprivation.

Hence, we see evidences everywhere of a cross-modal plasticity of the brain, allowing visual imagery without visual perception [36, 36, 37, 39, 40, 44, 46, 47]. The interesting point, at this moment, is how this plasticity works and how can we deal with it to provide novel approaches to mental images induction. The activation of PVAs is not homogeneous among blind subgroups, and it is specifically the case of the primary visual cortex (the so-called V1). In [48], a trans-age study is performed, regarding the activation of these areas within different groups of sighted, early and late blind people. Figure 1.6 presents the differences in brain activation for these three groups.

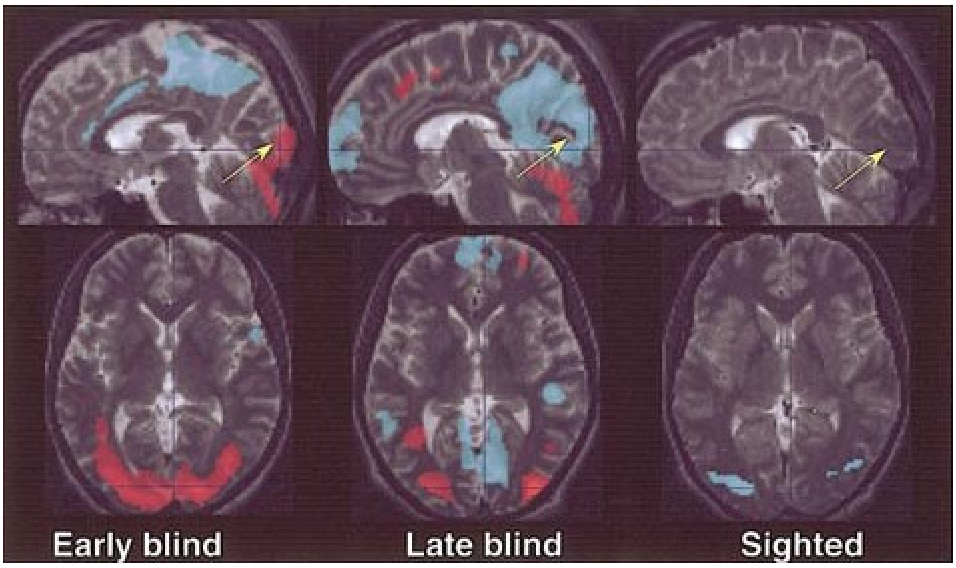


Fig. 1.6. Mental activity of early, late blind and sighted people [48].

This flexibility of the brain is called the cross-modal plasticity [38]. There have been reported results showing three types of cross-modal plasticity [38]:

- Early sensory areas of spared modalities.
- Primary cortices associated with the deprived modality.
- Polymodal association areas.

This can be done by means of different processes [38]:

- Changes in local connectivity.
- Changes in subcortical connectivity.
- Changes in cortico-cortical feedback.
- Changes in long-range cortico-cortical connectivity.

There is, hence, a dependency of the age, since some cross-modal changes can only occur during some steps of the development.

In [48], the authors suggest that 16 years-old is the limit for a functional shift of the V1 from processing visual stimuli to tactile stimuli in congenital blind people.

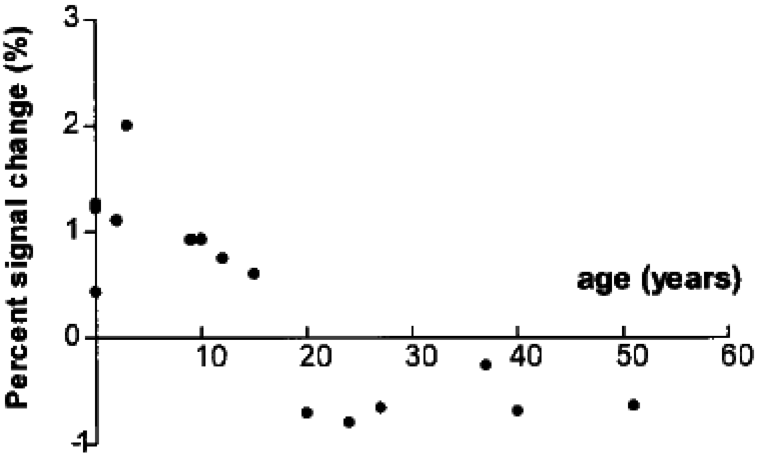


Fig. 1.7. Mental activity of the V1 area in 15 blind subjects [48].

We can find some psychological studies reinforcing the importance of providing O&M instructions as part of blind children rehabilitation curricula [16, 49-51].

There are two antagonist reasons dealing with this plasticity:

- Regarding this last study [48], early blind and young people are more suitable to reconfigure their brain structures, both primary and multimodal areas, than older people.
- Regarding [52], this process have some difficulties when the primary visual areas have been already recruited by other mental functions.

Finally, we can find measurements of these changes not only by means of RMI or PET, but by electro-encephalographs (EEG) [53], which is much cheaper and affordable for our study. This possibility opens the door to objective evaluations of the global AP designed.

1.5. Scope and Goal of Presented Work

Every O&M devices, following [16, 54], must:

- Establish and maintain a straight line of travel,
- Successfully perform and recognize changes in direction,
- Circumvent objects in the path of travel while maintaining a basic line of travel, and
- Recover from veers and other unintended or unexpected changes in direction.

We will see in the State of the Art that some problems are still to be overcome in APs for mobility.

The motivation section has shown some possibilities that are still to be explored, since the so-called cross-modal plasticity can reach to important reorganization of the brain structures which allow the brain to "see".

The global hypotheses are:

- H1: Technology can help visually impaired people to move securely.

- H2: Light and fast image processing algorithms can be designed, and can run over cheap and generic hardware.
- H3: Sounds can substitute mobility relevant aspects of the visual world (the so-called s*onification*).
- H4: Commercial technology is mature enough to implement a functional and low-cost AP.

The aim of this work is focused on the design of a mobility system for assisting visually impaired people, using image processing and sonification techniques to detect and represent potential obstacles in indoor and outdoor environments. Although specific research objectives will be proposed and discussed in detail in chapter 4, we will give a summary of them in the following lines.

To achieve the main objectives it is necessary to accomplish the following partial objectives (which will be detailed in chapter 4):

- Objective 1: Image processing: The image processing will be the sub-system responsible of retrieving the most mobility relevant information of the visual world. This is a crucial aspect of the system, regarding speed, accuracy, computational load and sensors. The main objectives of this sub-system will be high performance, low error rate (to assure safeness), low computational load and low cost of hardware needed.
- Objective 2: Sonification: The device will use sounds to transmit the surroundings and other mobility relevant information. The set of sounds must be clear, intuitive and easy to be learnt. They will be transmitted leaving free the ear for other real-world sounds, also important for a safe travel.
- Objective 3: Hardware: Given that the system must help in mobility tasks, it must be portable, with light weight and autonomous enough to work for some hours. Thus, the hardware must be small, and low power. Likewise, and due to other reasons, the hardware must be as cheap as possible, to help in the spreading of the AP.

Finally, a user study participated by visually impaired volunteers and experts must the proposed system showing the effectiveness of the solution implemented.

With all this summarized information, we can define the following partial goal:

- Price: The price shows to be a very important parameter regarding the final implementation of any AP. To develop a cheap device is a must if we want this study to be more than, precisely, a study.
- Usability: Simplification of use, simplification of information, provided with intuitive methods and channels. Moreover, the system must be light in terms of weight, with autonomy enough, and reliable to avoid dangerous crashes or accidents.
- Short training: Following the previous objectives, a short and easy training program must be designed. The training uses to be the hardest barrier for many people (specially aged or with special cognitive limitations). Thus, how do they approach the device for the first time must be carefully defined to avoid *a priori* rejections.

1.6. Chapters Outline

This work is divided in the following chapters and issues:

1. Introduction: The present chapter introduces the assistive products, blindness, the mobility and orientation concepts and the motivation and objectives of the present work.
2. State of the Art: In this chapter there is a review of the already proposed ETAs, specially focused on those similar to that presented in this study.
3. Users' Requirements: Summarization of the interviews performed to design and asset some specific problems regarding usability, training, etc.
4. Proposal: The design of the proposed ETA is explained in deep in this chapter. Specific objectives are also proposed in each research area such as image processing, sonification or hardware implementations.
5. Image Processing: Since the image processing is an important part of this work, a specific chapter is dedicated to this field, reviewing existing algorithms and presenting a new proposal. This chapter is closed with a comparison among algorithms and a final conclusion section.
6. Sonification: The way we can transform images into sounds is explained in this chapter. A review of some psychoacoustic issues and found proposals for sonification are also presented, as well as mine. A discussion is also provided.
7. Global Integration and Validation: Both integrated software and hardware projects are presented in this chapter, with the corresponding validations by users.
8. Conclusions, Further Works and Contributions: Although each chapter presents its own conclusions, global conclusions about the work are provided in this chapter, as well as a suggestion of some research lines in the field of assistive products for mobility for the blind. Also, articles, conference communications and other peer-reviewed published paper about this study are summarized in this final chapter.
9. Annexes: Some relevant data about the study is available in the annexes.

The main research areas are, thus, image processing, sonification and hardware implementation. They are not completely isolated, as shown in figure 1.8.

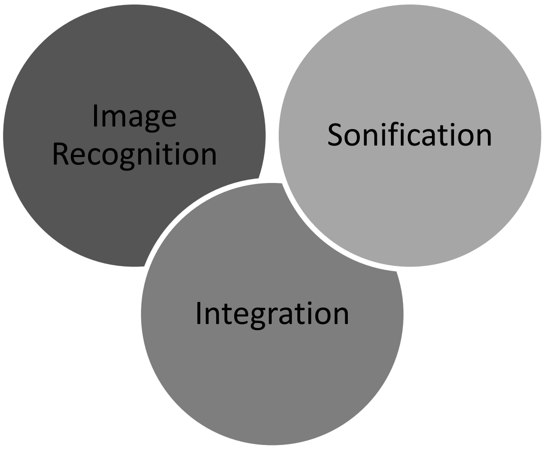


Fig. 1.8. Research areas of the thesis.

Each chapter presents its own references list in order to help the consultation of such references, if needed.

References

1. AENOR, "Productos de Apoyo para personas con discapacidad. Clasificación y Terminología (ISO 9999:2007)", 2007.

2. Y. F. Heerkens, T. Bougie, and M. W. d. K. Vrankrijker, "Classification and terminology of assistive products." *International Encyclopedia of Rehabilitation*. JH Stone and M Blouin, eds. *Center for International Rehabilitation Research Information and Exchange (CIRRIE)* . 2010.

3. A. Ocklitz, "Artificial respiration with technical aids already 5000 years ago?" *Anaesthesist* vol. 45 no. 1, pp. 19-21. 1996.

4. P. Revuelta Sanz, B. Ruiz Mezcua, and J. M. Sánchez Pena, "Sonification as a Social Right Implementation." *18th International Conference on Auditory Display*. *Proceedings of the 18th International Conference on Auditory Display (ICAD 2012* , pp. 199-201. 2012. Atlanta, GA.

5. R. M. Mahoney, "Robotic Products for Rehabilitation: Status and Strategy." *Proceedings of ICORR '97: International Conference on Rehabilitation Robotics* , pp. 12-22. 1997.

6. D. L. Edyburn, "Rethinking Assistive Technology." *Special Education Technology Practice* vol. 5 no. 4, pp. 16-23. 2004.

7. R. E. Ladner, "Accessible Technology and Models of Disability." *Design and Use of Assistive Technology: Social, Technical, Ethical, and Economic Challenges*. M.M.K.Oishi et al.(eds.), ed. no. 3, pp. 25-31. 2010. Springer Science+Business Media, LLC.

8. WHO, "Change the Definition of Blindness." . 2009.

9. WHO, "Blindness." *http://www.who.int/topics/blindness/en/index.html* . 2009.

10. WHO, "Visual impairment and blindness." *http://www.who.int/entity/mediacentre/factsheets/en/* . 2009.

11. WHO, "Maps." *http://www.who.int/blindness/data_maps/en/index.html* . 2009.

12. WHO, S. Resnikoff, D. Pascolini et al., "Global data on visual impairment in the year 2002." , pp. 844-851. 2004.

13. WHO, "Causes of blindness and visual impairment." *http://www.who.int/blindness/causes/en/index.html* . 2009.

14. F. Mast and T. Zaehle, "Spatial reference frames used in mental imagery tasks. Blindness and brain plasticity in navigation and object perception." New York: Lawrence Erlbaum Associates, ed. 2008.

15. R. Long and E. Hill, "Establishing and maintaining orientation for mobility." *Foundations of orientation and mobility*. In Blasch BB, Wiener WR, and Welsh RL, eds. vol. 2nd. 1997. New York: American Foundation for the Blind.

16. S. J. La Grow, "Orientation to Place." Center for International Rehabilitation Research Information and Exchange (CIRRIE), ed. vol. International Encyclopedia of Rehabilitation, pp. 1-8. 2010.

17. E. Hill and P. Ponder, "Orientation and mobility techniques: A guide for the practitioner." New York: American Foundation for the Blind., ed. *Rehabilitation Literature* vol. 38 no. 9, pp. 297-298. 1977.

18. W. Jacobson, *The art and science of teaching orientation and mobility to persons with visual impairments*: New York: American Foundation for the Blind., 1993.

19. S. J. La Grow and M. Weessies, "Orientation and mobility: Techniques for independence." . 1994. Palmerston North, New Zealand: Dunmore Press.

20. C. Martinez, "Orientation and Mobility Training: The Way to Go." Texas Deafblind Outreach. 1998.

21. H. Petrie, V. Johnson, V. Strothotte et al., "MoBIC: An aid to increase the independent mobility of blind travellers." *British Journal of Visual Impairment* vol. 15, pp. 63-66. 1997.

22. J. Reiser, "Theory and issues in research on blindness and brain plasticity." *Blindness and brain plasticity in navigation and object perception*. In: Rieser JJ, Ashmead DH, Ebner FF et al., eds. 2008. New York: Lawrence Erlbaum Associates.

23. T. Carroll, "Blindness: What it is, what it does, and how to live with it." B. Boston: Little, ed. 1961.

24. B. Lownfeld, "Effects of blindness on the cognitive functioning of children." *Nervous Child* vol. 7, pp. 45-54. 1948.

25. D. A. Ross and B. B. Blasch, "Wearable Interfaces for Orientation and Wayfinding." *ASSETS'00* , pp. 193-200. 2000.

26. J. Gill, "Priorities for Technological Research for Visually Impaired People." *Visual Impairment Research* vol. 7, pp. 59-61. 2005.

27. E. Trefler and D. Hobson, "Assistive Technology." C.Christiansen & C.Baum, ed. vol. Occupational therapy: Enabling function and well-being (2nd Ed.) no. 20, pp. 483-506. 1997.

28. R. Smith, "Technological Approaches to Performance Enhancement." In C.Christiansen & C.Baum (Eds.), ed. vol. Occupational Therapy: Overcoming Human Performance Deficits, pp. 474-788. 1991. Thorofare, NJ.

29. A. Cook and S. Hussey, "A Framework for Assistive Technologies." In A.Cook & S.Hussey (Eds.), ed. vol. Assistive Technology: Principles and Practices, pp. 45-76. 1995. St. James.

30. G. E. Hinton and L. M. Parsons, "Frames of Reference and Mental Imagery." J. Long and A. Baddeley, eds. no. 15, pp. 261-277. 1981.

31. H. Ehrlichman and J. Barrett, "Right hemispheric specialization for mental imagery: a review of the evidence." *Brain Cogn.* vol. 2 no. 1, pp. 55-76. 1983.

32. M. W. O'Boyle and J. B. Hellige, "Hemispheric asymmetry, early visual processes, and serial memory comparison." *Brain Cogn.* vol. 1 no. 2, pp. 224-243. 1982.

33. J. Hochberg, "Contemporary Theory and Research in Visual Perception." R. N. Haber, ed. vol. In the mind's eye. 1968. New York.

34. A. Noë, "Is the Visual World a Grand Illusion?" *Journal of Consciousness Studies* vol. 9 no. 5-6, pp. 1-12. 2002.

35. S. E. Palmer, *Vision Science: Photons to Phenomenology,* Cambridge,MA: MIT Press, 1999.

36. A. Ishai, "Seeing faces and objects with the "mind's eye"," *Archives Italiennes de Biologie,* vol. 148, no. 1. pp.1-9, 2010.

37. M. Behrmann, G. Winocur, and M. Moscovitch, "Dissociation Between Mental-Imagery and Object Recognition in A Brain-Damaged Patient," *Nature,* vol. 359, no. 6396. pp.636-637, 1992.

38. D. Bavelier and H. J. Neville, "Cross-modal plasticity: where and how?" *Nature Reviews Neuroscience* vol. 3 no. 443, p.452. 2002.

39. E. Mellet, N. Tzourio, F. Crivello et al., "Functional anatomy of spatial mental imagery generated from verbal instructions," *Journal of Neuroscience,* vol. 16, no. 20. pp.6504-6512, 1996.

40. A. G. De Volder, H. Toyama, Y. Kimura et al., "Auditory triggered mental imagery of shape involves visual association areas in early blind humans," *Neuroimage,* vol. 14, no. 1. pp.129-139, 2001.

41. C. Büchel, C. Price, R. S. J. Frackowiak et al., "Different activation patterns in the visual cortex of late and congenitally blind subjects," *Brain,* vol. 121. pp.409-419, 1998.

42. A. Amedi, N. Raz, P. Pianka et al., "Early 'visual' cortex activation correlates with superior verbal memory performance in the blind," *Nature Neuroscience,* vol. 6, no. 7. pp.758-766, 2003.

43. H. Burton, R. J. Sinclair, and D. G. McLaren, "Cortical Activity to Vibrotactile Stimulation: An fMRI Study in Blind and Sighted Individuals." *Human Brain Mapping* vol. 23, pp. 210-228. 2004.

44. B. Boroojerdi, K. O. Bushara, B. Corwell et al., "Enhanced excitability of the human visual cortex induced by short-term light deprivation," *Cerebral Cortex,* vol. 10, no. 5. pp.529-534, 2000.

45. A. Pascual-Leone and R. Hamilton, "The metamodal organization of the brain," *Vision: from Neurons to Cognition,* vol. 134. pp.427-445, 2001.

46. H. Bértolo, "Visual imagery without visual perception?" *Psicológica* vol. 26, pp. 173-188. 2005.

47. R. L. Gregory, "Seeing after blindness," *Nature Neuroscience,* vol. 6, no. 9. pp.909-910, 2003.

48. N. Sadati, T. Okada, M. Honda et al., "Critical Period for Cross-Modal Plasticity in Blind Humans: A Functional MRI Study." *Neuroimage* vol. 16 no. 2, pp. 389-400. 2002.

49. R. Crouse and M. Bina, "The administration of orientation and mobility programs for children and adults." *Foundations of orientation and mobility*. W. W. W. R. e. In Blasch BB, ed. vol. 2nd. 1997. New York: American Foundation for the Blind.

50. P. Hatlen, "The core curriculum for blind and visually impaired students, including those with additional disabilities." *http://www.tsbvi.edu/instructional-resources/1211-the-core-curriculum-for-blind-and-visually-impaired-students-including-those-with-additional-disabilities* vol. 28 no. 1, pp. 25-32. 2007. 14-4-2011.

51. S. Lewis and C. Allman, "Educational programming." *Foundations of Education*. K. A. e. In: Holbrook MC, ed. vol. 2nd. 2000. New York: American Foundation for the Blind.

52. J. A. Ferrero Blanco, "Imágenes Acústicas y Visión Inducida: Entrevista a don José Luis González Mora." *U.T.L.A.I.Punto Doc* . 2002.

53. J. Li, Y. Tang, L. Zhou et al., "EEG dynamics reflects the partial and holistic effects in mental imagery generation." *Journal of Zhejiang University-SCIENCE B (Biomedicine & Biotechnology* vol. 11 no. 12, p.-944. 2010. 951.

54. D. Guth and J. Reiser, "Perception and control of locomotion by blind and visually impaired pedestrians." *Foundations of orientation and mobility*. W. W. W. R. e. In Blasch BB, ed. vol. 2nd, pp. 9-38. 1997. New York: American Foundation for the Blind.

2. State of the Art

In the present chapter, we will present a review of already designed and, in some cases, commercialized assistive products. As it will be shown, this research field started long time ago, and then we find also important to give a historical overview of these first steps in applying technology to help disabled people. Moreover, we will see how the idea of some of the first projects has had a great success, and became a research line that is still open.

After that, a proposal of taxonomy is also provided, to order the huge amount of technical aids. However, this order will not always be followed, since the variability and flexibility of some APs make them hard to categorize in this taxonomy.

The review of mobility and orientation APs focuses, specially, on those similar to that presented in this work. This is mandatory to manage the basic ideas lying under these designs, as well as to generate critics, which may help us to develop more usable APs.

Finally, we find interesting to overview examples of the rest of them.

2.1. Brief History of APs for Orientation and Mobility

Several technologies and tools have been used since the beginning of humankind, such as canes. Some of the most advanced bets in the application of technology to improve blinds' life were the prototypes of Noiszewski's Elektroftalm (1897) and D'Albe's Exploring Optophone (1912) [1]. The technology limitations of this period were evident, and a democratized use of technical aids had to wait untill the 40's. First ETAs appeared in the U.S.A., such as the Signal Corps in 1943 [1], the Sensory Aid in 1948 [2], the Optar in 1950 [1]. In the next decade, and with the financial support of the Army and Veterans Administration, the Haverfort Colledge built the G-5 Obstacle detector [1, 3]:

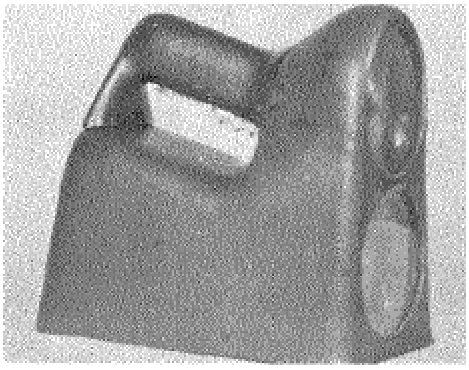


Fig. 2.1. The G-5 Obstacle detector, from [3].

As stated in [3], 1964 marked the beginning of the second decade of ETAs, because the laser had recently appeared. The G-5 rapidly became the Laser Cane, which used to cost, in the early 70's, around 1,900$ [3]:

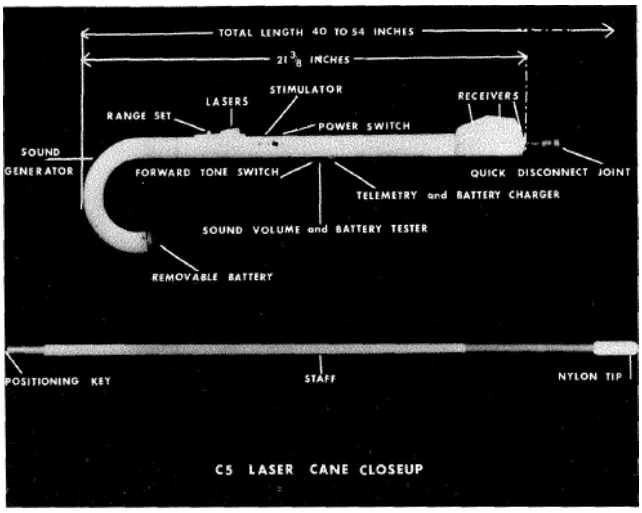


Fig. 2.2, the C-5 Laser Cane, from [3].

The idea presented in the Laser Cane, as we will see, has been widely exploited in many ETAs, so it is important to present it more in detail. As shown in figure 2.3 [1], the main idea of this first generation of ETAs was to produce some energy (light, but we will find also ultrasounds) which reflects in the bodies in front of the user. This reflection is received by photodiodes or other kind of receivers and the system produces the alarm signal.

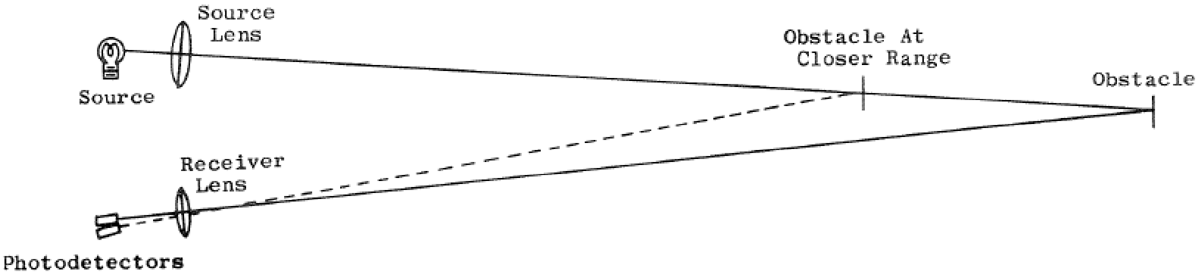


Fig. 2.3. Basic scheme of the Laser cane and that of most of the active ETAs, from [3].

This idea was also applied in some other devices developed in the 60’s and 70’s, such as the Ultrasonic torch and the ‘K’ sonar, both in 1965 [4-6], the Pathsounder in 1966 [1], the Torch in 1969 [1], the Mowat sensor in 1970 [1, 7, 8], and others. Table 2.1 summarizes the tools presented in these years.

40’s	50’s	60’s	70’s
Signal Corps [1]	Optar [1]	G-5 [1]	Mowat Sensor [8]
Sensory Aid [2]		C-5 (Laser Cane) [1]	The FOA Sawdish Laser cane [1]
		Ultrasonic Torch [6]	Mims IR Mob. Aid [1]
		‘K’ sonar [4]	Sonic Glasses [7]
		Pathsounder [1]	Nottingham Obstacle Detector [1]
		Torch [1]	Swedish Laser cane [9]
			Single Object Sensor [1]
			Tactile Vision Substitution System [1]

Table 2.1. Brief summary of ETAs developed from the 40’ till the 80’.

2.2. Taxonomy of Orientation and Mobility APs

Classification of EOAs and ETAs is not a simple task. These products can be seen from many different points of view.

The first division of ICTs applied to blindness mobility and orientation is, obviously, the difference between mobility and orientation. Although in [10] this division is done in three parts (ETAs, EOAs and Position Locator Devices), for better presentation purposes we can assume that every AP can fit in some of the two basic categories, ETAs and EOAs, including this last group the position locator devices. After this first and basic partition, both ETAs and EOAs can be classified attending to the following parameters:

Parameter	Space of possibilities
Energy Management	Active
	Passive
Technology	Ultrasounds
	Incandescent light
	Infrared
	Laser
	GPS
	Compass
	Mono-Stereo Vision
	RFID
	WLAN
	Gyroscope
	GSM/GPRS/UMTS (Mobile Phone Network)
	Bluetooth

	Inertial sensors
Hardware use	Belt
	Cane
	Chest
	Hand
	Head Mounted
	Neck
	Phone
	Shoe
	Skin
	Tonge
	Weared
	Earphones
	External (implemented over the urban space)
Information channel	Tactile Electro Mechanic
	Vibration
	Sounds (unstructured) (stereo or mono)
	Synthetic Voice
	Recorded Voice
	Braille
	Bone conduction
	Mechanical guidance
	Direct Stimulation of Brain Cortex
Information structure	Discrete (from binary to 'N' symbols)
	1 dimension
	2 dimensions
	3 dimensions
	More dimensions or combinations
Field of application	Indoor
	Outdoor

Table 2.2. Meta-classification of ETAs and EOAs.

This preliminary meta-classification is not self-exclusive; the hardware use is closely related to the information channel. In the same way, the technology used determines if the system is active or passive, the field of use, the accuracy, etc. Moreover, there are some APs which, for example, use different technologies, or have different user profiles providing information in different ways, by different channels, or implemented over the user and the environment.

We propose in this chapter the following classification hierarchy (hierarchy level represented by numbers), when applicable:

1. Electronic Travel Aids (ETAs)
    2. External/Carried by the user/Mixed
        3. Hardware use
        3. Information channel and structure

1. Electronic Orientation Aids (EOAs)
    2. Indoor/Outdoor/Mixed
        3. Technology
    2. Information channel

1. Mixed EO&TAs

2.3. Related Devices and Proposals

In this section we will present a review of the technology available for ETAs and EOAs implementation, followed by a state of the art of developed APs following the proposed taxonomy. We find interesting to deep in details with some paradigmatic APs in each classification field because, as we will see, some of them are quite similar. At the end of this section, the analysis of some devices hardly classifiable in this taxonomy will be carried out.

2.3.1 Technology Review

Most of the proposed APs to help blind people in mobility and orientation are based on a small set of technologies, most of them commercially available, what has helped to develop cheap and useful devices.

1. Incandescent Light: The first option in the history was based on incandescent light. The emission and reflection of this light was used to roughly detect obstacles, with the scheme shown in figure 2.3. This technology showed a lot of problems, for example, the day light which interacts with the emitted light.

2. Infrared (IR): Infrared light is the part of the electromagnetic spectrum around 900 and 1600nm wavelength. This part of the spectrum is invisible for us, but emitters and receptors are quite cheap, so they have experimented a massive use in APs and, specifically, in sensing systems. The basic scheme was already presented in figure 2.3. From the user point of view, there is no difference between ultrasounds and infrared based devices.

3. LASER: Laser is the acronym of “Light Amplification by Stimulated Emission of Radiation”, a technology that produces a highly correlated source of light, which can be used for several applications, among them, distance measurements.

4. Ultrasounds: These devices have one or more emitters of high-frequency sounds (from 40KHz to some hundreds of KHz) and the corresponding sensors. Figure 2.4. shows the working principle:

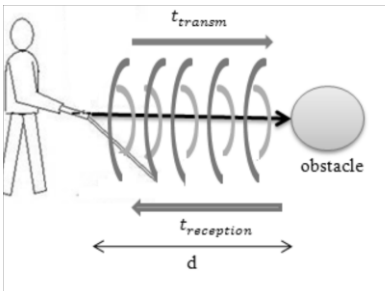


Figure 2.4. Example of ultrasound sensing, the eCane, from [8].

Measuring the Time-of-Flight (TOF), $t_{transm}+t_{reception}$ in the figure, these devices obtain the distance to an object.

5. RFID: Most modern technology has been proposed to help visually impaired people to move and get oriented. The Radio Frequency Identification (RFID) generates a short distance communication between a base station and a tag, sensitive to the electromagnetic field generated by the base, and taking its power from it. This system has shown to be very cheap and, in some situations, useful.
6. GPS: The Global Positioning System is a worldwide satellites network which allows calculating the position of a receptor comparing several synchronized signals from these satellites. As it will be shown, this technology started to be usable in the 90's, and has been embedded in several devices. However, it presents some important drawbacks, like the deficient functioning in indoors or the underground.
7. Compass, gyroscopes and Inertial sensors: Autonomous and self-references have also been used to help the blinds to get oriented. Compasses use the magnetic natural field of the earth to calculate directions regarding the north. Gyroscopes can give information about inclinations or directions. Finally, the inertial sensors measure changes in directions and speeds, so the system can calculate where the user is pointing, for example. The general limitation of these systems is the lack of a global reference (except the case of the compass, and in this case, very limited).
8. Image (Stereo or mono): Visually processing the surroundings of the user has been another possibility, overall in the last years, given the decrease of the price of cameras and micro-computers. This option tries to emulate the brain work towards mobility information, and may do it processing mono or stereo images to get the information of the environment. Again, no global reference is provided by this approach.
9. Wireless Networks (WLAN, Bluetooth, GSM…): Nowadays we can find several wireless networks surrounding us. WiFi networks, Cell phones networks, Bluetooth (BT), etc. can be used to get oriented. Moreover, these technologies are widely spread and used, and the prices are very low. However, they present some limitations. In the case of wide networks, such as WiFi or cell phone networks, the accuracy is not high enough for mobility tasks (around 50m error). BT presents the opposite problem: very small areas so it cannot be used to help the user to get oriented in spaces wider than a couple of meters. Finally, none of them can be automatically adapted to changes in the environments (like a new scaffold in the street), and must be carefully and

manually modified (putting the transceivers in the correct place to signal them). In other words, they need important infrastructures.

2.3.2 Mobility APs (ETAs)

We have defined mobility as the skills which make the subject “able to carry out a plan to get there”, i.e., to a specific place [11]. These are the so called Electronic Travel Aids (ETAs).

2.3.2.1 External ETAs

Few researches have been performed regarding the installation of different devices in the urban space, as well as in indoor environments, prosecuting mobility purposes. The main problem of this approach is the cost and very specific task of the system. Mobility devices have to take into account “micronavigation” aspects, as road works, so they should be placed to cover all the possible paths and provide real-time feedback. Moreover, the border line between mobility and orientation when dealing with this family of ETAs is not clearly defined. Micronavigation meets macronavigation when identifying obstacles, in a pre-defined environment, helps to orientate the user. A proposal that we find closer to mobility than to orientation is the Walking and Sport Support system [12]. This system uses cameras (and image processing techniques) placed, for example, along the speedway, and they help the blind user to run safely. The main problem was the hardware needed for a safe run, and the system has never been applied.

2.3.2.2 ETAs Carried by the User

In this class of APs we find most of the ETAs developed. Avoiding obstacles and providing a safe travel depends on the specific movements of the user, so “global” solutions are hard to be designed and implemented. On the other hand, user-carried ETAs fit much better for this kind of problem. In the historical review we found that the first ICTs applied to blindness were included in this group (and so it happens with non-ICT technology such as the canes).

2.3.2.2.1 Torch-like ETAs

Torch-like ETAs were some of the first prototypes; more specifically, ultrasounds based ETAs. We have previously seen some of these examples, such as the G-5 Obstacle detector or the Signal Corps (even if limitations in technology forced to build “bag-like” ETAs). After those devices, Kay developed, in the late 60’s, the Ultrasonic torch and the Torch, setting up the paradigm of this approach [1, 6, 7]. After them, the Mowat Sensor [8], the Tactile Handle [13] and the Polaron [14] appeared. The most modern implementation is the Miniguide [15]. Figure 2.5 presents the Miniguide:

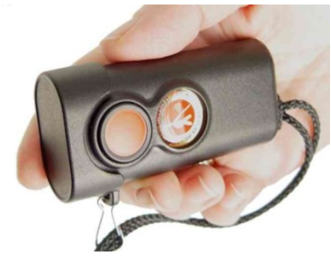


Fig. 2.5. The miniguide, from [15].

However, some APs torch-like and IR based can be also found, as the case of the Hand Guide [8], shown in figure 2.6.

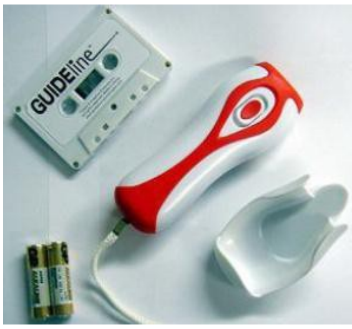


Fig. 2.6. The Hand Guide [8].

Finally, we present the University of California Santa Cruz (UCSC) Project [16], which is a handled ETA, laser based. The most important limitation of this ETA, is that it has not been tested with blind people and, thus, its application is not clear enough yet. As it is shown in figure 2.7, its size is quite big.

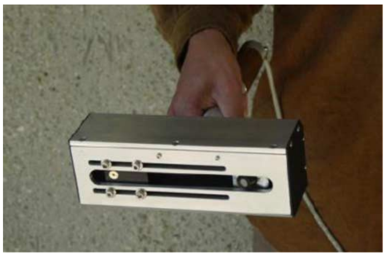


Fig. 2.7. The UCSC project, from [10].

Finally, an interesting option in this family with IR implementations uses far IR, which means "thermal" vision [17]. The main advantage of this last implementation is that the system is passive, which means it emits no energy, receiving the naturally emitted radiation by warm bodies.

2.3.2.2.2 Cane-like ETAs

The problem with the torch-like devices is that one hand of the user is employed to handle the ETA. Moreover, blind users hardly renounce to the white cane, since it is the most reliable AP to prevent from falls. Then, both arms and hands are occupied by different tools. Because of that, cane-like products started to appear, being the laser cane the first of them. The main idea was presented in figure 2.3, regardless we deal with ultrasonic, IR, laser or incandescent light devices. Examples of ultrasonic devices in this category are the 'K' sonar [4, 5], the Ultracane [18], the eCane [8], the Digital Ecosystem Sytem [19], the Electric White Cane [20], the Dynamic Ultrasonic Ranging System [21], or the Walk Mate [8]. As said before, these implementations have the advantage of employing only one hand, indeed, in the same manner that the white cane, as shown in figure 2.8.

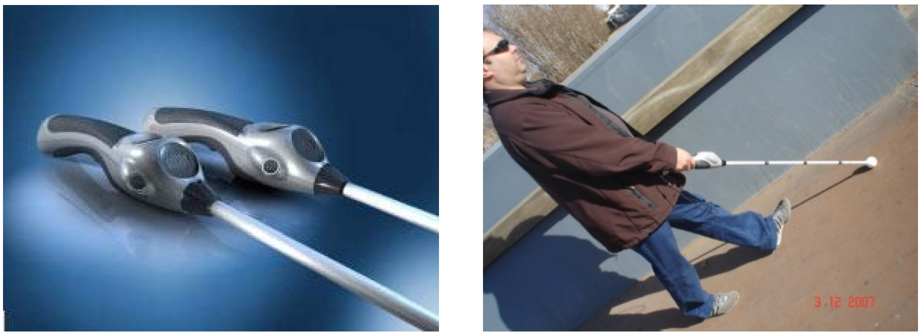


Fig. 2.8. The Ultracane [18].

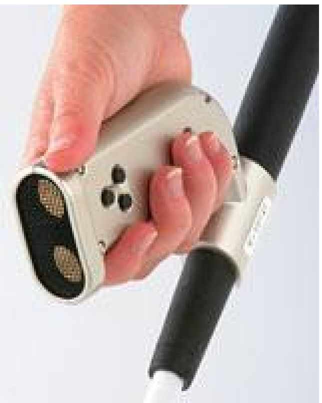


Fig. 2.9. The 'K' sonar [8].

An interesting prototype, proposed by Ulrich and Borestein [22] is the Guide Cane. This robot perceives by means of ultrasounds sensors the free path and mechanically guides the user:

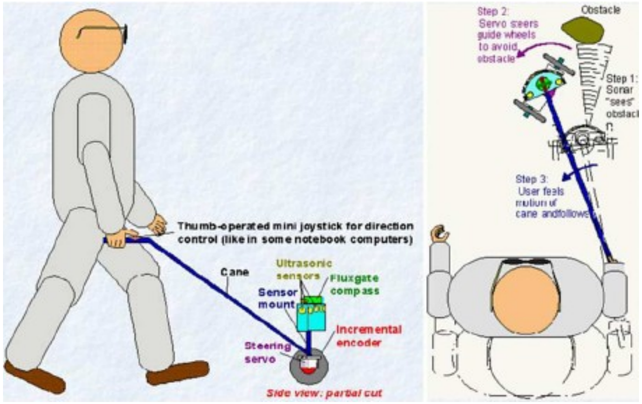

Thumb-operated mini joystick for direction control (like in some notebook computers)
Cane
Ultrasonic sensors
Fluxgate compass
Sensor mount
Incremental encoder
Steering servo
Side view: partial cut
Step 2: Servo steers guide wheels to avoid obstacle
Obstacle
Step 1: Sonar "sees" obstacl n
Step 3: User feels motion of cane andfollows

Figure 2.10. The Guide cane [10].

This proposal is, however, expensive and not so useful in irregular pavement environments, such as rural streets.

Likewise, most of the IR based ETAs are designed to be mounted on the cane. This is the case of Tom Pouce [23], working at 950nm wavelength, which is one of the simplest implementations among these solutions:

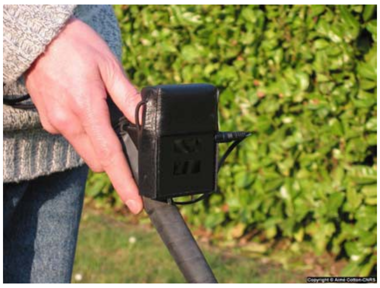


Fig. 2.11. Tom Pouce [23].

The correlative IR based ETA to the Guide cane, is the PALM-AID [24], shown in figure 2.12.

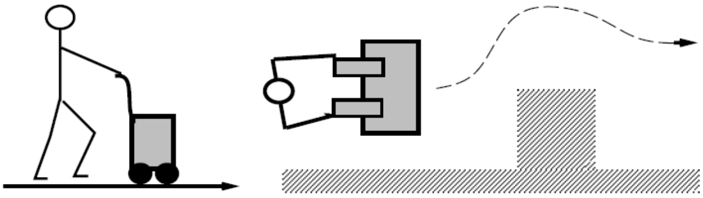


Fig. 2.12. PALM-AID [24].

Additionally, many other ETAs were previously reported and available in a commercial way. We find, for example, since 1972 the FOA Swedish Laser cane [1, 9, 25], the Teletact already in the 90's [23], the Laser Long cane [8] or the Laser Orientation Aid for Visually Impaired (LOAVI) [26]. This last device is shown in the next figure.

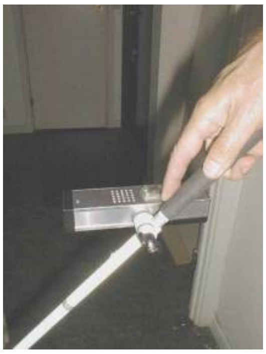


Fig. 2.13. The LOAVI assistive product [26].

2.3.2.2.3 Belt or Wearied ETAs

As stated in [10], the main problem of many mobility devices is the occupation of the hands to be able to use them. This is applicable for both torch and cane-like ETAs, even if this last solution is handier for the user. However, some other implementations have been proposed. Being the first one the Navigational Aid for the Blind [27], working with two infrared emitters, the most known and ultrasound based ETA in this group is the so called NavBelt [28].

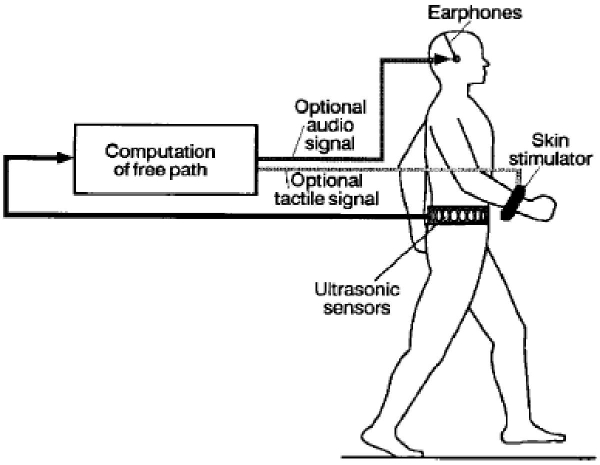


Fig. 2.14. The NavBelt [28].

Sensors are mounted in a belt, while the computation unit is carried in a bag. Other belt based ETAs can be found in [29-33]. This last project, developed in the MIT, is shown in the next figure.

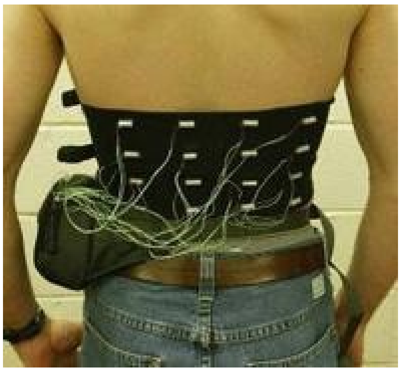


Fig. 2.15. MIT belt project [34].

Another implementations hang from the neck and use other information channels, like the Guelp Project "Haptic Glove" [35-37]. This device, based on stereo vision processing, holds from the neck of the user, as shown in figure 2.16. The data are sent to the user by means of tactile gloves.

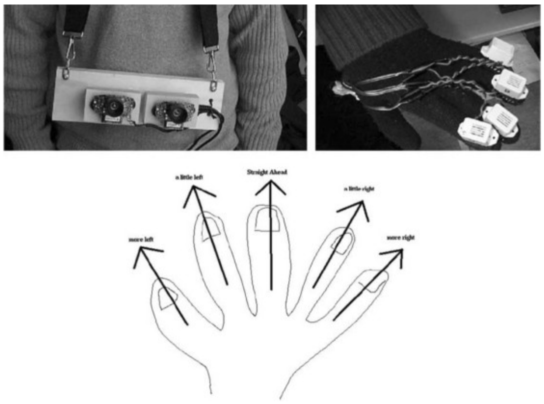


Fig. 2.16. The Guelp Project (Haptic Glove) [10].

In order to make the ETAs less ostensive, researchers have tried other possibilities like the chest [8, 14], the tongue [38, 39]or they are simply wearied in the vests [19, 31, 40], with very small implementations. An example of this idea is the EPFL Project [10, 41]:

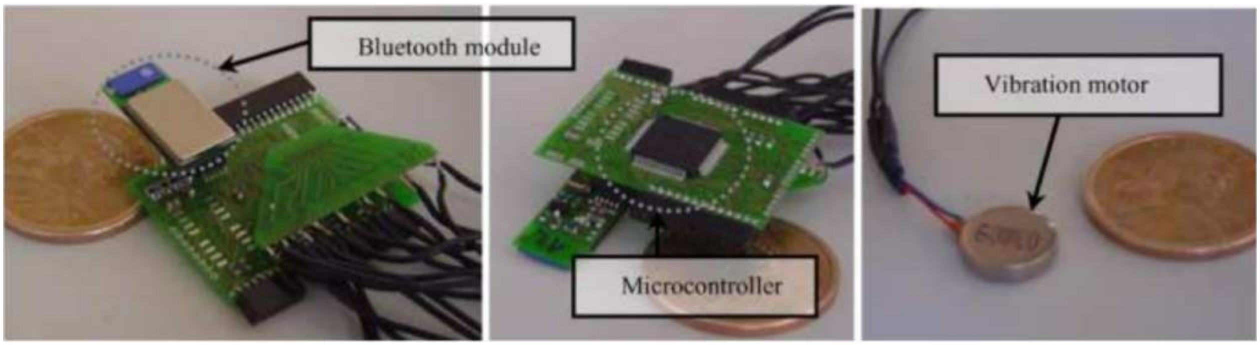


Fig. 2.17. EPFL project hardware [10].

An example of wearied sensors and actuators can be seen in the next figure.

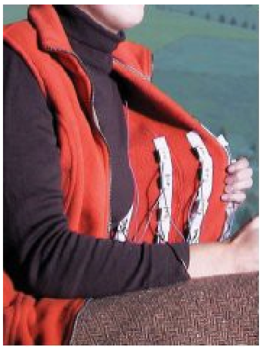


Fig. 2.18. TNO Project [34].

Finally, figure 2.19 shows a shoe implementation from [42]:

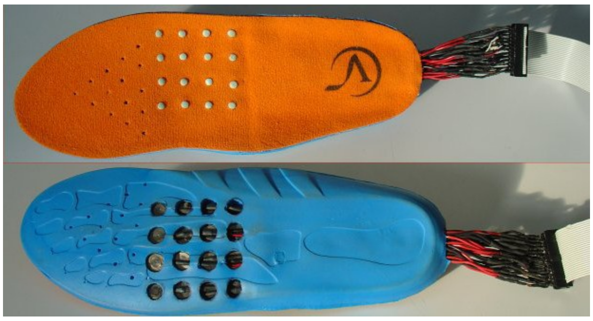


Fig. 2.19. Shoe tactile display for the blind [34].

2.3.2.2.4 Head Mounted ETAs

We must study in a separate section the head mounted ETAs, because of two main reasons: On one hand, because most of them propose a vision substitution, and not only an obstacle avoidance function. On the other hand, because our proposal follows this paradigm, and it is interesting to study them in detail. Following this way of simplifying the use of devices, a set of APs that are head mounted have been proposed using different technologies. This idea was firstly proposed by Kay, in 1966, with the Binaural Sonic Aid (also known as Sonicguide [1, 5, 43, 44], working with ultrasounds. This device is shown in figure 2.20.

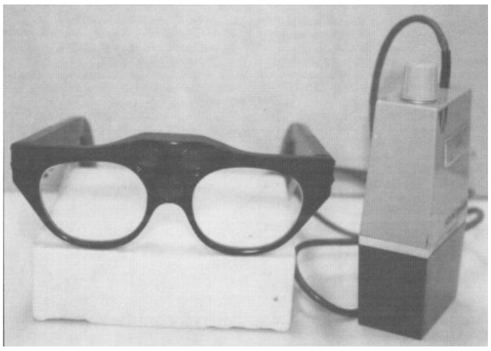


Fig. 2.20. The Sonicguide, from [43].

Other devices in this research line are the Sonic Pathfinder [45-47] or the Sensory 6 [48]. Most modern implementations are the KASPA [49] (which is the continuation of the Sonic Glasses), the binaural sonar [50], the Wearable Collision Warning System [51] and others [52, 53]. All these head mounted devices work with ultrasound emitters and sensors. The first example of a head mounted and IR based ETA are the Mims IR Glasses, developed in 1972 and shown in figure 2.21.

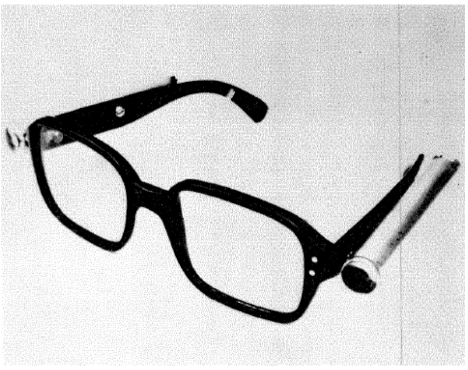


Fig. 2.21. Mims IR glasses [1].

A practical device could be implemented byIR glasses are also based on IR, achieving obstacle detection even in darkness [54].

A more complex possibility to inform the user about his or her surroundings is to shortcut the ways with which the brain "sees", in order to produce mental images by means of other sensory and cognitive paths [55-57], such as verbal instructions [58, 59], sounds [60, 61] or tact [62, 63]. The so-called cross-modal plasticity is exploited by many ETAs, which have proved to provide mental imagery to the trained users. An example of this approach is an IR based device, which provides a complex representation of the environment with sounds: The depth support for the vOICe project (the project itself will be presented in detail further in this section).

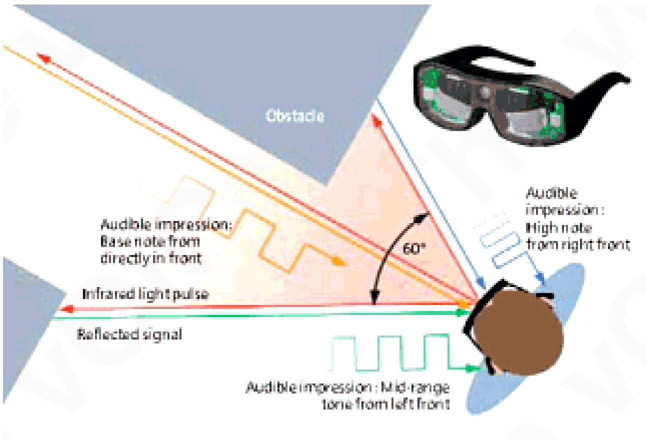

Obstacle
Audible impression: Base note from directly in front
60°
Audible impression : High note from right front
Infrared light pulse
Reflected signal
Audible impression : Mid-range tone from left front

Fig. 2.22. The depth support for the vOICe [64].

Finally, we will present a laser based and head mounted tool, also providing acoustic maps to the user: the Computer Aided System for Blind People (CASBliP) [36], developed in 2008. The hardware is shown in the next figure.

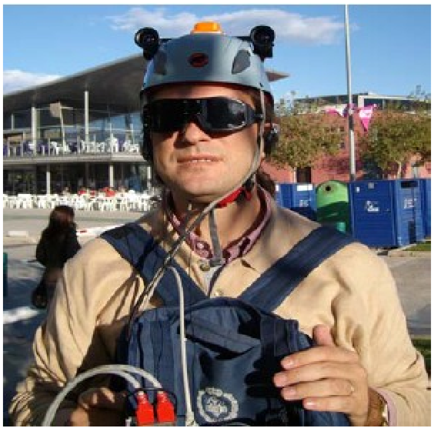


Fig. 2.23. The CASBliP Project [36].

Although the good results achieved by these first options of head mounted ETAs, we can see how the simplicity and usability for the user was not always achieved. Moreover, the decrease of computational costs in portable devices has opened the door to real time image processing. Moreover, as said before, image processing systems have an important advantage regarding the ultrasonic, (near) infrared or laser devices: they are passive. Less energy is needed, systems are cheaper and lighter.

When dealing with image processing, ETAs can be divided in two main categories:

- Monocular vision: One single camera is used to manage the environmental information.
- Stereo vision: Two cameras, slightly separated are used to compute depth maps and, hence, distances to objects.

In the first group, depth is not extracted and, hence, information about obstacles must be managed in different manners. Examples of these systems are monocular vision based ETAs that can be found in [29, 65-67] as well as in the following reviews [10, 14, 25, 34]. In the case of the Navigation Assistance for Visually Impaired (NAVI) [68], the system presents one camera which captures the scene in front of the user. A single-board processing system separates the fore and the background in order to represent, via stereo sounds, the objects in the aforementioned scene.

Special attention must be taken with a monocular vision (later extended to depth perception as explained before) system: the vOICe [69-71] (where "OIC" comes from the expression "Oh! I can *See*"). The hardware of this system is presented in the following figure.

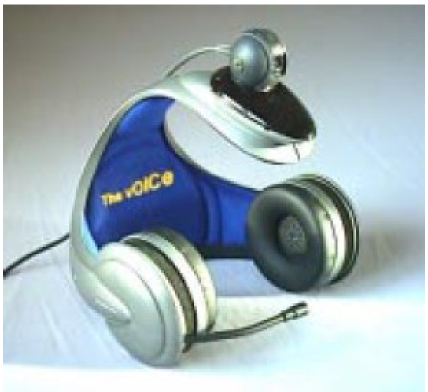


Fig. 2.24. vOICe helmet [70].

The system captures the scene, as explained in the NAVI system. In this proposal, the gray scale image is processed and converted to sounds following the next rules:

- Each pixel is converted to a sinusoidal wave.
- Horizontal position of the pixel is transformed to time.
- Vertical position is transformed to pitch of the wave.

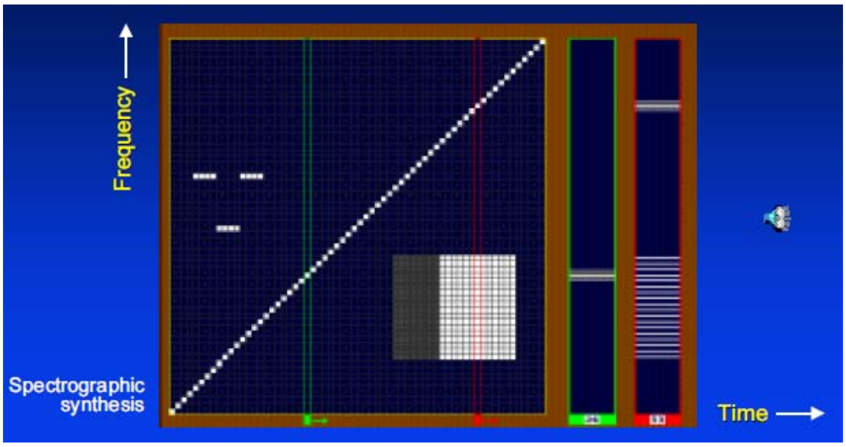


Fig. 2.25. Image-to-sound transform in the vOICe [70].

Users' feedbacks show that the system arrives to generate subconscious images of their surroundings, crossing from "darkness into light" [71]. Despite these approaches, the distance to the object is a relevant information for a safe travel. Thus, other systems were already designed for a depth perception and transformation into sounds. This is the case of the Virtual Acoustic Space [72], the SVETA system [73, 74], the Visual Support System for the Blind [75-77] or the Brigham project [78]. The main problem when codifying an image or a 3D scene into sound (the so called *sonification*) is the overload of the auditory system [25, 36]. Thus, there have been proposed other channels to provide the information to the user. We have already presented the Haptic Glove system ([35-37] and figure 2.16), although it was not a head mounted device. There are also head mounted devices avoiding the use of the hearing system, such as the Electro-Neural Vision System (ENVS) [79]:

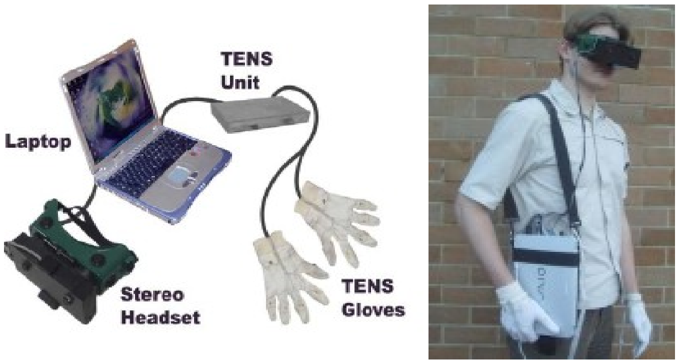


Fig. 2.26. ENVS [79].

This last system implements a haptic based stimulation, by means of a pair of gloves connected to a laptop, which process the 3D information from the surroundings.

A final alternative we are going to present is the Prof. Kahn project, head mounted and ultrasound based device with direct nerve stimulation at the wrist [80]. This stimulation presents some technical problems, as well as some risks. The frequency of the stimulus is inversely related with the distance, achieving up to 500Hz at 20cm, while amplitudes need to be around 100-200V to effectively stimulate the superficial nerves under the skin.

2.3.2.3 Carried and External ETAs

It is mandatory to present a mixed group of ETAs, which are carried by the users, but simultaneously they used pre-installed hardware in the environment. Most of them use radio waves to inform where the user is, and this information is transmitted to the user by means of verbal information or other methods. Again in this section, we find difficulties to separate the proposed APs between mobility and orientation. The paradigm of this family of APs is the Remote Guidance system [81].

The scheme of the Remote Guidance aid is shown in figure 2.27.

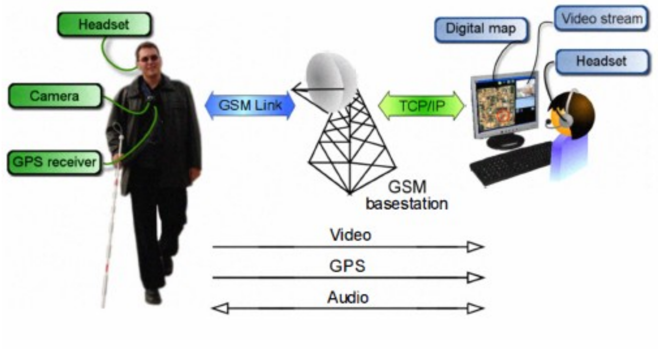
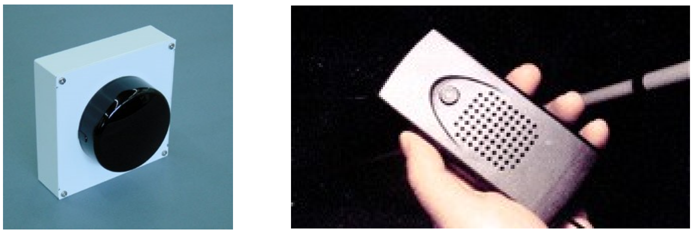


Fig. 2.27. Remote guidance scheme, from [81].

This system needs a human-controlled tele-center to guide the user and help him or her to avoid obstacles. Obviously, it implements a two-way information transmission, so it is important to have a network with wide coverage. Indeed, the implemented solution is based on the mobile phone networks. It provides the mobility information to the user by means of vocal messages. Other systems are the Metronaut [82] based on laser identification of tags with a cane or the Electronic Guide Stick [7], which interacts with sensors implanted in the curve, acting as a guide. The Smart Bat Cane [83] and the Vibrator Cane [84] use radio frequencies to identify some features of the external world (for example, a semaphore with an emitter identifying itself as semaphore). Additionally, the Smart Bat Cane also uses ultrasounds to help in mobility and obstacle avoidance. Both of them use vibrations to inform the user about distances and, in the case of the Smart Bat Cane, recorded voice is also transmitted to the user for more detailed information. We find some other proposals based on RFID. These are the cases of SeSaMoNet [85], the Robot-Assisted Indoor Navigation System [86], the Blind Interactive Guide System [87] or the RFID Navigation System [88].

Lastly, there is an intensively implemented, tested and commercialized system, which uses infrared beacons to locate and help the user in his/her mobility. This is the Remote Infrared Audible Signage (RIAS) system [89]. This system, funded by the Federal Transit Administration, U.S.A., was designed after the conclusions of some case studies with visually impaired people [90, 90], showing the difficulties of this collective regarding the use of the public transport.

The RIAS hardware consists of two devices: the first one is a transmitter and the second one a receiver as is shown in the following picture.



(a) (b)

Fig. 2.28. RIAS hardware: (a) permanently installed transmitters and (b) receivers [89].

The main idea of this system is the following:

1. The transmitter emits a verbal message coded in an IR beam.
2. The receiver transforms this IR message to audible sound.

The different components of the system interact as shown in figures 2.29 and 2.30. The commercial name of the RIAS is Talking Signs®, prosecuting the following goals [91]:

1. Using auditory signage in simple location and wayfinding tasks.
2. Remotely identifying which bus to take.
3. Selecting the correct bus in a congested mixed mode setting (i.e. multiple buses,
4. cars, vans, trucks).
5. User evaluation of auditory signage technology.

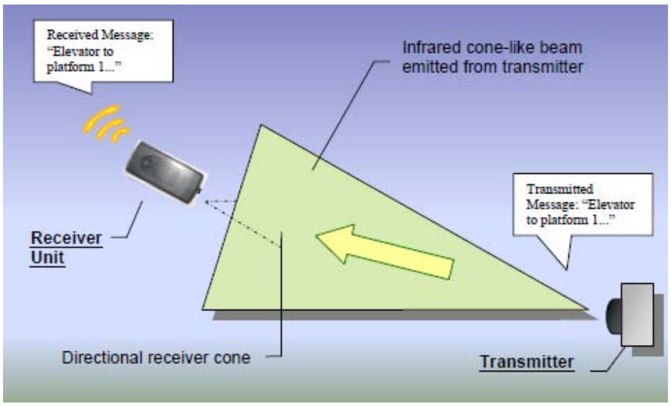

Received Message: "Elevator to platform 1..."
Infrared cone-like beam emitted from transmitter
Transmitted Message: "Elevator to platform 1..."
Receiver Unit
Directional receiver cone
Transmitter

Fig. 2.29. RIAS working scheme [89].

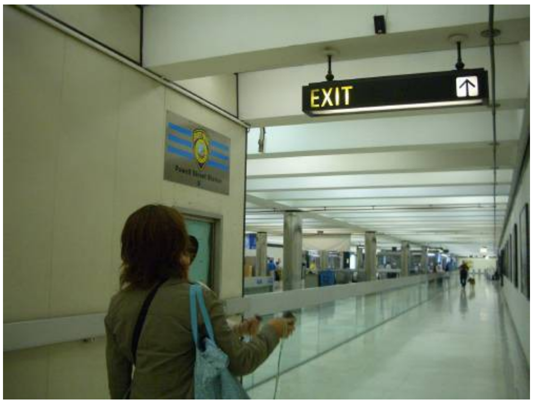


Fig. 2.30. Woman using RIAS in the Powel Street Station, in San Francisco [89].

2.3.3 Orientation (EOAs)

Orientation means to "know where he is in space and where he wants to go" [11]. Thus, a global position reference system is mandatory, as well as some kind of maps recorded in the devices. This proposals focus on different aspects of independent living and travel of blind people than those presented before. The location in a global space and not only in the immediate surrounding needs different strategies and technologies, but we will find, as well, some shared technologies between the two main approaches studied in this chapter.

Some general aspects can be discussed regarding orientation APs:

- The orientation information, given a map or a path, uses to be discrete; hence, most of the EOAs use synthetic or recorded voice to transmit verbal instructions about the position and direction of the user, with different paradigms. However, we can find some exceptions to this rule, in some EOAs that orientate the user through his path by alerting with unidimensional information when hi/she is leaving the correct path.
- Regarding the technology, the GPS is the most used system to locate and orient the user, since it is a well-established system, cheap and accurate enough to implement this kind of APs. Sadly, some problems will arise when orienting blind people in indoor environments. Some different technology is then proposed, such as radio frequency devices, inertial sensors, compasses, gyroscopes or even laser, infrared or mono vision systems. If the GPS and maps information is taken as external support for the EOAs, we will then avoid the classification regarding "external" and "autonomous" EOAs, since all of them are, following this criterion, external and carried by the user.

However, an "indoor-outdoor" classification is possible, since the environment of application force the technology used to implement each AP.

- These APs do not substitute the previously presented ones, the ETAs. Indeed, they are complementary. Because of that, we will find some mixed APs, which integrate orientation and mobility capabilities. These will be left till the end of this section.

2.3.3.1 Indoor EOAs

Indoor navigation systems don't use global positioning resources (such as the GPS). Because of that they appeared before than those using GPS and other global positioning technology. The way they obtain the references to locate and orientate the user is done by means of beacons. Examples of this early technology are the Uslan projects [92, 93], and those proposed by Lancioni [94, 95].

Two paradigms of this family of EOAs can be represented by the Blind Interactive Guide System (BIGS)[87] and the Mobile and Position Orientation System (PYOM) [96]. As "indoor" technologies, none of them use GPS or other coverage sensitive systems. Instead of that, the BIGS uses RFID tags placed in strategic locations to orientate the user. These tags only transmit their identification (ID), and this information is processed by a portable computer. No more infrastructures are needed and it is assumed that detailed maps with information about the beacons are available. The path is stored in the portable device. RFID has been proposed as beacon technologies by other proposals, such as those presented in [97, 98]. Following the same architecture of beacons, but proposing another technology to implement beacons and their communication with the user device, we find the "infrared" option. In this family we find, for example, the Cyber Crumbs [99]. Finally, proposals by Lancioni use radio frequency beacons to detect the user [94, 95, 100].Sometimes the portable device can be carried in the cane (or another devices), as we saw in the ETAs, as it is the case of the Instrumentation Cane [101], which helps the user, regarding the orientation task, by means of laser (reading printed tags), gyroscopes and a pedometer. With this information, the system is able to track the path followed by the user and, thus, orientate him/her in the correct direction. The architecture of these systems is composed by user' devices, coordinated with different kinds of beacons providing the orientation information to that device. The other paradigm is represented, among others, by the PYOM, which presents a server-client architecture in phase 1, as shown in figure 2.31. In phase 2, no server is needed and the location is computed by measuring intensities from different access points [96].

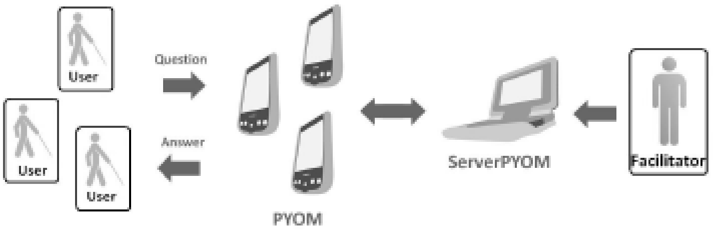


Fig. 2.31. Server and clients in phase 1 of PYOM [96].

Previously in the same way, Uslan proposed two systems for guidance by recorded voice and a network of loudspeakers [92, 93], using IR light to detect the position of the user. A server, again, selected the appropriated message to help the user navigating in the pre-installed environment. Both proposals achieve the same objectives, but implementing the complexity of the orientation in different places; the first model, needing quite complex portable devices, the second one, with very simple portable devices, but needing a server. As user's available technology has developed very fast in the last years, most of this kind of proposals, and many of them in the outdoor field are based on computationally complete and powerful portable devices.

There is a final proposal, which is hardly classifiable as indoor EOA (because t is just a software) is the Subway Mobility Assistance Tool [102]. This is a program that helps blind people to design their trip around the underground public transport system. As we will see in the next section, software and programs have been proposed, aiming to reduce the final price of these EOAs.

2.3.3.2 Outdoor EOAs

We can find outdoor orientation devices some years before global positioning systems were available for the public. Based on tactile maps, we find, for example, the NOMAD system [103], shown in figure 2.32.

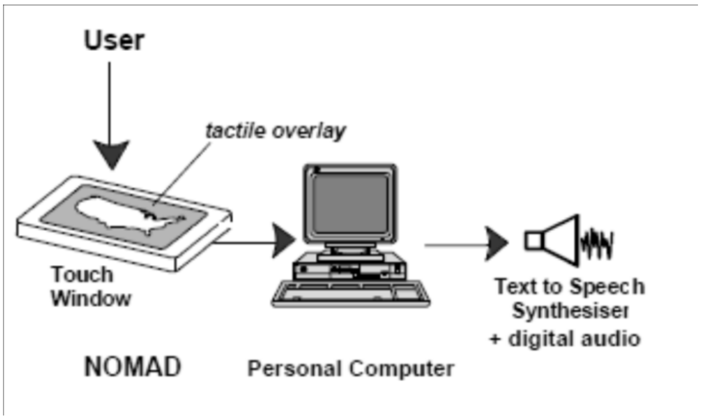


Fig. 2.32. NOMAD system [104].

In these years, no global positioning system was available and, hence, the orientation task was left to the user, providing the proposed systems just maps to help him/her to get oriented [105-107]. However, outdoor guidance and orientation systems have seen a huge growing period since the GPS technology achieved usable accuracy, in the 2000 year [108]. This technology created a new paradigm for orientation tools (and not only for disabled people). A global and objective reference was available and, hence, automatic and global orientation was possible. Before this year, some proposals were given using, however, the GPS system: the Personal Guidance System [105], the Dodson guide [109], the Strider [110], the Navigation System for the Blind [111] and the MoBIC [112, 113]. This last EOA was designed in two main components: The MoBIC Pre-Journey System (MoPS), "to assist users in planning journeys, and the MoBIC Outdoor System (MoODS) to execute these plans by providing users with orientation and navigation assistance during journeys" [112]. The main problem with these approaches was the huge amount of memory needed to store digitalized maps.

In the decade of the 2000’, we find an explosion of GPS based EOAs. Table 2.3 presents a summary of these proposals.

System	Year	Reference
BrailleNote GPS	2002	[114]
MobileGeo	2002	[115]
Victor Trekker	2003	[116]
Loadstone GPS	2004	[117]
Active Belt	2004	[118]
Wayfinder	2005	[119]
NOPPA	2006	[120]
Géotact	2006	[23]
Trinetra	2006	[121]
MOST-NNG	2008	[122]
Wayfinder Open Source	2010	[119]
Kapsys	2011	[123]

Table 2.3. Summary of some GPS based EOAs appeared since 2002.

Since all mentioned projects work in a similar manner, only one of them will be presented. The BrailleNote GPS is a fully accessible GPS device, providing the orienting information to the user by means of a Braille line and a speaker. It allows computing travel routes and other GPS based location information. The BrailleNote GPS appeared in 2002, without street maps or other points of interest. This information had to be stored by the user [108].The second version, in 2003, already incorporated that data. Figure 2.33 hows two photographs of the BrailleNote GPS.

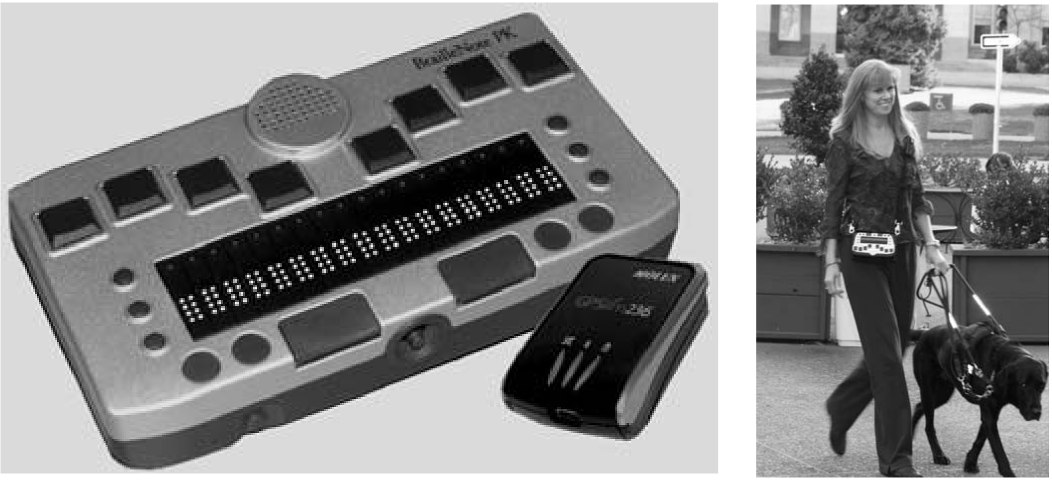


Fig. 2.33. The BrailleNote GPS, alone and being used by a blind person, from [108].

There is, however, an exception of outdoor EOAs which does not use GPS. The way the project of the University of Osnabrück achieves to obtain the correct direction is implemented by means of a compass [124]. This EOA is a belt-based, with different actuators that provide the correct direction via vibrotactile stimulation.

2.3.3.3 Mixed Indoor/Ourdoor EOAs

A final group of EOAs, designed to work in both indoor and outdoor environments (with the proper installation and maps), is presented in this section. These EOAs use to implement the orientation procedures by means of different technologies. Such a combination makes devices more complex, but also more flexible when an adaptation to different environments is needed. However, there is a first group of this family only implementing one technology. The Easy Walker [7, 125], the Blind Orientation System [125] or the Sound Buoys project [14]. These three systems use infrared beacons in both indoor and outdoor environments to obtain the user's location. Other EOAs, as said before, use a combination of technologies to achieve the proposed capabilities, being able to work in indoor and outdoor scenarios. This is the case of the NOPPA project [120] and the Indoor Navigation System [126], which uses GPS, but also Bluetooth, WIFI and other wireless technologies, and that of the Trinetra [121], with GPS, GSM and RFID to assist blind people in their shopping.

As it has been done in other sections, we would like to present a singular and technologically differenced project. The Body Mounted Vision System [127] is an orientation tool which uses pre-recorded paths to compare the current path followed by the user, by extracting some characteristic points and features of the surrounding. A monovision system captures images from the real path and a portable image-processing device searches some relevant points to obtain the position of the user. The image processing, then, matches the extracted points with the recorded ones, so the system knows where it is. The main problem of this implementation is that pre-recorded paths are needed, and any change in them may cause malfunctions or unexpected behaves.

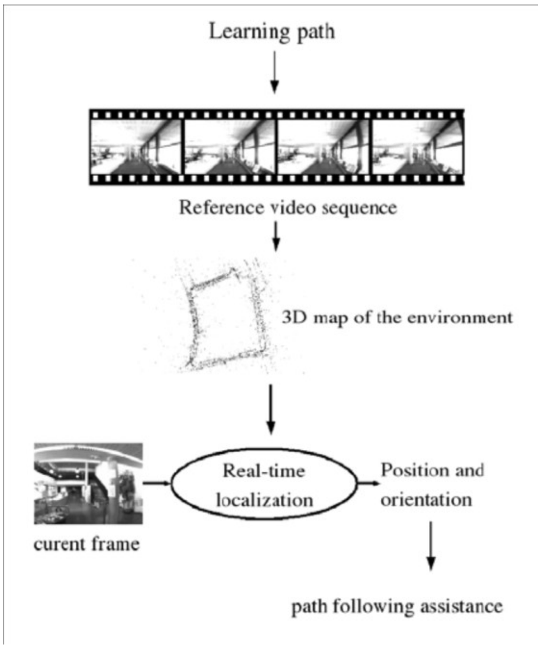


Fig. 2.34. The Body Mounted Vision System synoptic, from [127].

2.3.3.4 Information Transmitted by EOAs

In the references section we find different proposals about how to communicate to the user the orientation information. The most important paradigms are the following:

- Tactile
    - Vibration: The orientation information is provided by different actuators which give the user an idea of the correct path, as done in [124].
    - Braille: Another tactile possibility is to implement Braille interfaces and, hence, provide verbal instructions via this reading code. This option is implemented in the BrailleNote GPS [114], for example.
- Hearing
    - Spatial language: Spatial language is defined as user referred words or sentences as "ahead", "behind", "X meters away", etc. This mode has been widely implemented, and even commercial GPS systems use this paradigm. In the case of EOAs, for example that prototype presented in [94] implements this mode of transmitting information. Another possibility is what Loomis *et al.* called the "no compass" mode, which only provides information of left-right with the error measurement expressed in degrees [111].
    - Clock metaphor: The clock metaphor uses the idea of the clock, assuming the user being oriented to the "12" hour, to provide directional information:

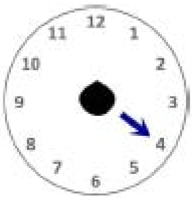


Fig. 2.35. Clock metaphor used in the PYOM project [96].

    This metaphor has proved to work quite well with blind users. The PYOM project implemented an analysis of the limitations of spatial language and finally decided to implement this option [96]. Another project in this way is the Géotact [23].

    - 3D sounds: As proposed in different ETAs, some proposals tried to orient the user by means of simulating the spatial position of sound sources. In [111] it is proposed the so called "bearing mode", where the position of the beacon or computed direction is represented by a synthetic voice, filtered by the head-related transfer function (HRTF), to help the user orient him/herself in the correct direction in a more natural manner. An example of this paradigm is the SWAN project [128].
    - Error sounds: A final paradigm is to alert the user when he/she looses the correct path. This option has been implemented, for example, in [94, 95, 127].

Some researchers have been carried out looking for difference in the performance of different ways to codify the spatial information. In this line, [129] found that 3D sound simulation helped the user to get oriented, in contrast to spatial language:

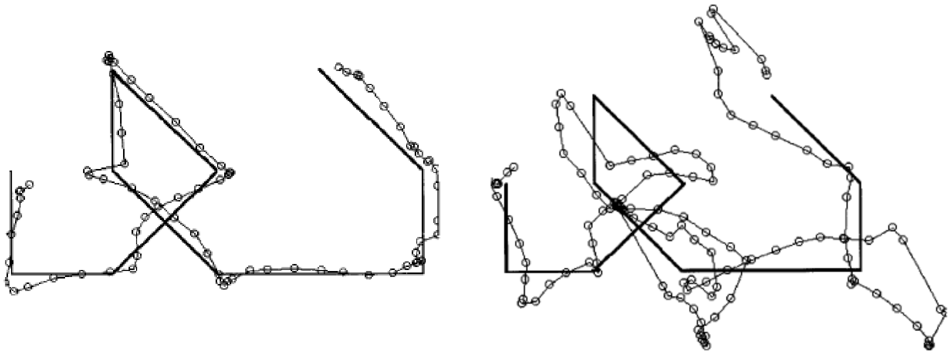


Fig. 2.36. The same path (straight lines) followed by blind people using "bearing mode" (left) and the "no compass mode" (right), which is much poorer [111].

The information provided to the user can be, however, simplified. In [127], we find a comparison of spatial language against a "bip" error alarm, whose frequency depends on the deviation angle:

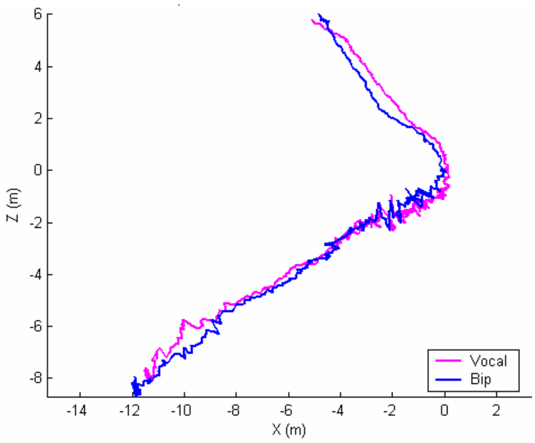


Fig. 2.37. Two orientation performances, with spatial language (in pink) and "bip" alarms (in blue), from [127].

We find, regarding these two studies, that the simpler and more natural is transmitted the information to the user, the more accurate and easier the directional decision is taken.

2.3.4 Orientation and Travel Aids (EO&TAs)

Not every project and commercial system available after these decades of research fits strictly in the EOA/ETA classification. Some of them have been designed to help in both displacement procedures. The SWAN project, for example, presents a mix of orientation technologies (such as GPS, beacons, etc.) with some mobility capabilities and sensors (like infrared). Thus, this system is able to provide orientation information at the same time of identifying the surrounding world, objects and textures [128]. Some of the most complete system proposed is the Navigation Assisted by Artificial Vision and GNSS (NAVIG) [130]. The scheme of this system is presented in figure 2.38.

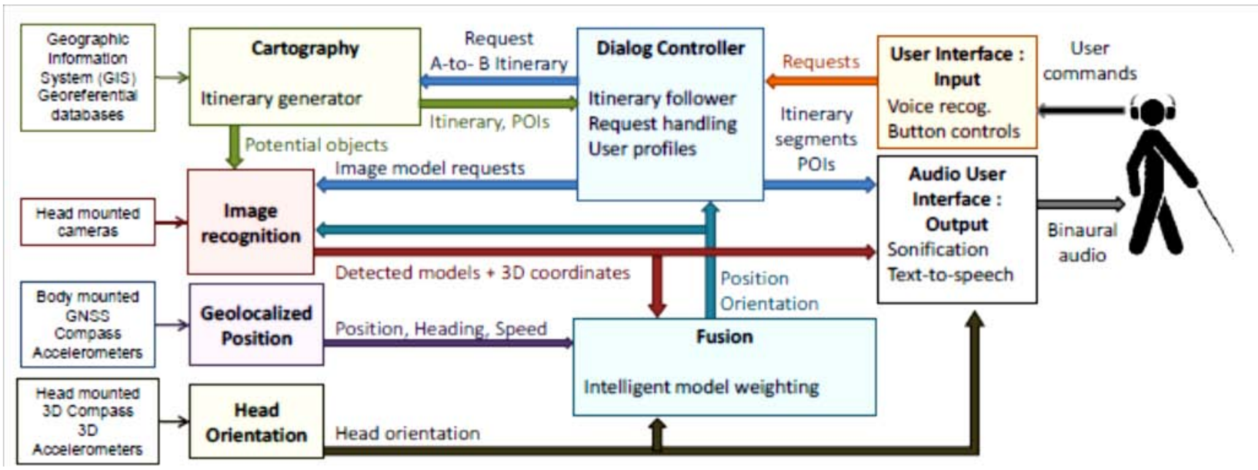


Fig. 2.38. NAVIG functional block, from [130].

We can easily identify the two main components of every ETA and EOA: a global reference system (in this case, the GIS, compass and the accelerometers) and a surrounding identification system (composed by head mounted cameras). All this information is integrated by some algorithm deciding which information to show the user and how. Another way to interpret the mix of orientation and mobility procedures can be focusing on static and dynamic data during the travel, as proposed in the Dishtri project [131]. The static one should refer to orientation and global directions, while the dynamic changes of the path (such as road works) should be treated and processed by different functional blocks in the same AP.

Final examples we would like to provide of this combination of technologies and approaches are the already presented Electro-neural vision system (ENVS) [79] and the Intelligent Assistant [132], which combines 3D vision processing with GPS (and in the case of the ENVS, a compass) to help in orientation and mobility.

2.4. Discussion

Assistive products have shown to be very versatile in the different application fields in the daily life of blind people. However, the final utilization of them has also shown to be variable. Thus, many projects are left in the cabinets of the universities. The some problems have been identified which block the access to the users:

- Commercial availability: Most of the APs presented in this study are lab prototypes and, hence, no price is given. Few of them have become commercial several years ago, but with high prices, as the laser cane (many years ago with a price of $ 1,900). Nowadays technology allows much cheaper implementations. Although we couldn't find prices to give a reference.
- The usability: When dealing with technology, and more concretely with electronics, designers and engineers may forget that the final user is the average person, not especially interested in technology. If the system requires a hard training period, if the information is not very intuitive or the control of the system is complex, users have shown to reject this kind of proposals. This effect is changing with the new generations but, for instance, many blind people are elderly and were not grown up with technology.

- Discreteness: Blind people don’t like, like the rest of people, to be forced to carry an especially ostentatious and showy system just to walk. We found in this review few discrete systems, which don’t point to be used by the user in public spaces.

The counterpart of these problems is the absolute success of the white cane, whose functionality, price and simplicity should act as guidelines for designers of any assistive product. Regarding the discreteness, the white cane has become a symbol (an international symbol, indeed), so it may not the paradigm to follow when designing new APs for the blind in this aspect.

In general terms, the use of commercial hardware, integrated circuits, sensors or even devices such as smartphones, as proposed by many EOAs should be extended to provide low-cost, light weight and miniaturized systems which really could help the blinds in their travels.

Moreover, different profiles should be included when designing new assistive products. People have different skills, attitudes and capabilities to understand the received information. Thus, every assistive product should be able to present information with different complexities, being a choice of the final user which degree of “detail” wants to receive. In many cases, the user may prefer to receive a “boolean” alarm instead of a complex combination of sounds, vibrations, etc.

Given that the hearing system is the most important sensing system for blind people, two antagonist research lines appear:

- Exploiting the hearing system: In this line we found the most complex proposals, in terms of information, since the hearing system presents a high sensitivity to many different aspects of sound waves. However, most of them presented a serious problem: they block the normal usage of this sense (by means of earphones, for example). Thus blind people often reject these systems. Therefore, other systems should be explored, like the bone transmission and other pseudo-hearing technologies. Psychoacoustics and HRTF processing’s appear to have achieved the asymptote as informational sources. Arbitrary but simple proposals in this line should be, proposed and tested in the short time.
- Exploiting other sensory paths: The other option has been to look forward other senses. This path is completely opened, especially regarding nerve or even cortical stimulation by means of different technologies. This path has to respect, however, the “integrity” of the user, and many of them will not agree to enter in the OR to “see” some neurophosphone effects. More non-intrusive approaches dealing with skin and nerve perception are still to be developed.

2.5. Conclusions

The assistive products design is a very active research field since the 70’s, and especially the last decade, because some technologies have become popular, cheap and powerful (like GPS, image sensors or smartphones). Thus, an explosion of these proposals can be found in the last years.

Additionally, many different technologies are involved in helping blind people to move around unknown or variable environments: RFID, wireless networks, infrared or laser light, ultrasounds, image processing, satellites, etc. Each technology has shown to have an application field, and outside this field it is no longer applicable. Because of that, some generalized APs, designed to be applied in many different environments, must incorporate different technologies and must also take the right decisions about the use of each technology in every moment.

However, some important problems were found, regarding the implantation of the proposals, focusing on usability and availability as the most important parameters to take into account for a viability study. We should add to this list the price. Even if we will not compare the final price with other ETAs, this is an important constraint because of obvious reasons.

Finally, we can find open research lines to be explored, aiming to find new communication and informational channels and paradigms, as well as smaller and cheaper devices using commercial hardware in order to build practical Assistive Products.

References

1. L. W. Farmer, "Mobility devices," *Bull Prosthet Res*. pp.47-118, 1978.

2. L. Cranberg, "Sensory Aid for the Blind." *Electronics* no. Mar., p.116. 1946.

3. J. M. Benjamin, "The laser cane," *Bull Prosthet Res*. pp.443-450, 1974.

4. M. T. Terlau, "'K' SONAR and Student Miniguide: Background, Features, Demonstrations and Applications." *CSUN Conference on Technology and Persons with Disabilities.Northridge, CA.* 2005.

5. L. Kay, "A Sonar Aid to Enhance Spatial Perception of the Blind: Engineering Design and Evaluation." *The Radio and Electronic Engineer* vol. 44 no. 11, pp. 605-627. 1974.

6. L. H. Riley, G. M. Weil, and A. Y. Cohen, "Evaluation of the Sonic Mobility Aid." vol. American Center for Research in Blindness and Rehabilitation, pp. 125-170. 1966.

7. M. Conti Pereira, "Sistema de Subtituiçao Sensorial para Auxílio a Deficientes Visuais via Técnicas de Processamento de Imagens e Estimulaçao Cutânea,", Sao Paulo, 2006.

8. A. R. Cardoso Costa, "e-Cane,", University of Minho, Portugal, 2009.

9. G. Jansson, "The effect of the range of laser cane on detection of objects by the blind." U. o. U. S. Dep.of Psychol., ed. vol. Report No. 211. 1975.

10. D. Dakopoulos and G. Bourbakis, "Wearable Obstacle Avoidance Electronic Travel Aids for Blind: A Survey." *IEEE Trans. on Systems, Man, and Cybernetics-PART C: Applications and Reviews* vol. 40 no. 1, pp. 25-35. 2010.

11. C. Martinez, "Orientation and Mobility Training: The Way to Go." vol. Texas Deafblind Outreach. 1998.

12. J. Batllé, A. B. Martínez, and J. Forest, "Specialized Colour Image Processor for Assisting Blind People in Walking and Sport Activities." *IMC'8 conference Proceedings* . 1996.

13. M. Bouzit, A. Chaibi, K. J. de Laurentis et al., "Tactile feedback navigation handle for the visually impaired." *Proceedings of IMECE20042004 ASME International Mechanical Engineering Congress and RD&D ExpoNovember 13-19, 2004, Anaheim, California USA* , pp. 13-19. 2004.

14. M. A. Hersh and M. Johnson, "Mobility: An Overview." *Assistive Technology for Visually Impaired and Blind People*. Marion A.Hersh and Michael A.Johnson, eds. no. 5, pp. 167-208. 2008.

15. GDP Research, "The Miniguide ultrasonic mobility aid." *http://www.gdp-research.com.au/minig_1.htm* . 2011.

16. D. Yuan and R. Manduchi, "A tool for range sensing and environment discovery for the blind." *Proc.2004 Conf.Comput.Vis.Pattern Recogn.* vol. 3, p.39. 2004.

17. D. S. Hedin, G. F. Seifert, G. Dagnelie et al., "Thermal Imaging Aid for the Blind." *Proceedings of the 28th IEEE, EMBS Annual International Conference, New York City, USA, Aug 30-Sept 3* , pp. 4131-4134. 2006.

18. M. C. Cruz Pedraza, "Informe de Valoración de Nuevos Productos: Ultracane." ONCE, ed. 2007. Lleida, Spain.

19. D. J. Calder, "Travel Aids For The Blind - The Digital Ecosystem Solution," *2009 7Th IEEE International Conference on Industrial Informatics, Vols 1 and 2*. pp.149-154, 2009.

20. N. Ñiacasha Utreras and N. Sotomayor, "Desarrollo de un Dispositivo que Mida la Distancia a un Objeto Emulando el Efecto de un Bastón Blanco para las Personas Invidentes,", 2004.

21. R. Gao, X and L. Chuan, "A dynamic ultrasonic ranging system as a mobility aid for the blind," *1995 IEEE Engineering in Medicine and Biology 17th Annual Conference and 21 Canadian Medical and Biological Engineering Conference (Cat.No.95CH35746)|1995 IEEE Engineering in Medicine and Biology 17th Annual Conference and 21 Canadian Medical and Biologi*. pp.10, 1997.

22. I. Ulrich and J. Borestein, "The GuideCane — Applying Mobile Robot Technologies to Assist the Visually Impaired." *IEEE Transactions on Systems, Man, and Cybernetics, -Part A: Systems and Humans* vol. 31 no. 2, pp. 131-136. 2001.

23. R. Farcy, R. Leroux, A. Jucha et al., "Electronic Travel Aids and Electronic Orientation Aids for Blind People: Technical, Rehabilitation and Everyday Life Points of View." *Conference & Workshop on Assistive Technologies for People with Vision & Hearing Impairments Technology for Inclusion CVHI 2006*. 2006.

24. K. M. Dawson-Howe and D. Vernon, "Personal Adaptive Mobility Aid for the Inrm and Elderly Blind." vol. Technical Report TR-CS-95-18. 1995. Computer Science Dept., School of Engineering, Trinity College Dublin.

25. A. Mittal and S. Sofat, "Sensors and Displays for Electronic Travel Aids: A Survey." *International Journal of Image Processing* vol. 5 no. 1, pp. 1-14. 2010.

26. S. Löfving, "Extending the Cane Range Using Laser Technique." *IMC9 conference Proceedings* . 2009.

27. R. Gangadharan, "Navigational Aid for the Blind," *Engineering in Medicine and Biology Society, 1990., Proceedings of the Twelfth Annual International Conference of the IEEE,* vol. 1-4 Nov. pp.2283, 1990.

28. S. Shoval, J. Borestein, and Y. Koren, "Auditory Guidance with the Navbelt-A Computerized Travel Aid for the Blind." *IEEE Transactions on Systems, Man, and Cybernetics* vol. 28 no. 3, pp. 459-467. 1998.

29. L. A. Johnson and C. M. Higgins, "A navigation aid for the blind using tactile-visual sensory substitution," *2006 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vols 1-15*. pp.869-872, 2006.

30. F. Gemperle, N. Ota, and D. Siewiorek, "Design of a wearable tactile display," *Fifth International Symposium on Wearable Computers, Proceedings*. pp.5-12, 2001.

31. H. van Veen and J. van Erp, "Providing directional information with tactile torso displays." *Proc.of EuroHaptics 2003* , pp. 471-474. 2003.

32. L. A. Jones, B. Lockyer, and E. Piateski, "Tactile display and vibrotactile pattern recognition on the torso," *Advanced Robotics,* vol. 20, no. 12. pp.1359-1374, 2006.

33. K. Ito, M. Okamono, J. Akita et al., "CyARM: An alternative aid device for blind person." *Proc.CHI05* , pp. 1483-1488. 2005.

34. R. Velazquez, "Wearable Assistive Devices for the Blind." *Wearable and Autonomous Biomedical Devices and Systems for Smart Environment: Issues and Characterization, LNEE 75*. A.Lay-Ekuakille and S.C.Mukhopadhyay, eds. vol. Springer no. 17, pp. 331-349. 2010.

35. R. Audette, J. Balthazaar, C. Dunk et al., "A stereo-vision system for the visually impaired." vol. Tech. Rep. 2000-41x-1. 2000. Sch. Eng., Univ. Guelph, Guelph, ON, Canada.

36. N. Ortigosa Araque, L. Dunai, F. Rossetti et al., "Sound Map Generation for a Prototype Blind Mobility System Using Multiple Sensors." *ABLETECH 08 Conference* , p.10. 2008.

37. J. Zelek, S. Bromley, D. Aamar et al., "A haptic glove as tactile vision sensory subsitution for way finding." *Journal of Visual Impairment & Blindness* , pp. 621-632. 2003.

38. P. Bach-y-Rita, K. Kaczmarek, M. Tyler et al., "From perception with a 49-point electrotactile stimulus array on the tongue: a technical note." *Journal of Rehabilitation Research and Development* vol. 35 no. 4, pp. 427-430. 1998.

39. M. Ptito, S. Moesgaard, A. Gjedde et al., "Cross-modal plasticity revealed by electrotactile stimulation of the tongue in the congenitally blind." *Brain* vol. 128, pp. 606-614. 2005.

40. S. Ram and J. Sharf, "The people sensor: a mobility aid for the visually impaired," *Digest of Papers.Second International Symposium on Wearable Computers*

(Cat.No.98EX215)|Digest of Papers.Second International Symposium on Wearable Computers (Cat.No.98EX215). pp.10, 1998.

41. S. Cardin, D. Thalmann, and F. Vexo, "A wearable system for mobility improvement of visually impaired people." *Vision Computing* vol. 23 no. 2, pp. 109-118. 2007.

42. R. Velazquez, O. Bazan, and M. Magana, "A shoe-integrated tactile display for directional navigation," *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2009)*. pp.1235-1240, 2009.

43. J. A. Brabyn, "Mobility Aids for the Blind." *Engineering in Medicine and Biology Magazine* vol. 29 no. 4, pp. 36-38. 1982.

44. N. C. Darling, G. L. Goodrich, and J. K. Wiley, "A preliminary followup study of electronic travel aid users," *Bull Prosthet Res,* vol. 10, no. 27. pp.82-91, 1977.

45. A. D. Heyes, "The use of musical scales to represent distance to object in an electronic travel aid for the blind." *Perceptual and Motor Skills* vol. 51 no. 2, pp. 68-75. 1981.

46. A. D. Heyes, "The Sonic Pathfinder - A new travel aid for the blind." *In Technology aids for the disabled*. W.J.Perk and Ed. s, eds. pp. 165-171. 1983. Butterworth.

47. A. D. Heyes and G. Clarcke, "The role of training in the use of the Sonic Pathfinder." *Proceedings of the American Association for the Education and rehabilitation of the Blind and Visually Impaired, Southwest Regional Conference, Hawaii.* 1991.

48. G. T. Campbell and J. C. Swail, "Sensory 6: Ultrasonic Mobility Aid for the Blind." *The Fourth International Conference on Mobility and Transport for the Elderly and Disabled Persons* , pp. 955-970. 1986.

49. L. Kay, "KASPA." http://www.batforblind.co.nz/ . 2005.

50. 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.

51. B. Jameson and R. Manduchi, "Watch Your Head: A Wearable Collision Warning System for the Blind," *2010 IEEE Sensors*. pp.1922-1927, 2010.

52. T. Ifukube, T. Sasaki, and C. Peng, "A Blind Mobility Aid Modeled After Echolocation of Bats," *IEEE Transactions on Biomedical Engineering,* vol. 38, no. 5. pp.461-465, 1991.

53. N. Debnath, J. Thangiah, S. Pararasaingam et al., "A mobility aid for the blind with discrete distance indicator and hanging object detection," *Tencon 2004.2004 IEEE Region 10 Conference (IEEE Cat.No.04Ch37582)*. pp.664-667, 2004.

54. L. Matthies and A. Rankin, "Negative obstacle detection by thermal signature," *Iros 2003: Proceedings of the 2003 IEEE/Rsj International Conference on Intelligent Robots and Systems, Vols 1-4*. pp.906-913, 2003.

55. H. Bértolo, "Visual imagery without visual perception?" *Psicológica* vol. 26, pp. 173-188. 2005.

56. R. L. Gregory, "Seeing after blindness," *Nature Neuroscience,* vol. 6, no. 9. pp.909-910, 2003.

57. D. Bavelier and H. J. Neville, "Cross-modal plasticity: where and how?" *Nature Reviews Neuroscience* vol. 3 no. 443, p.452. 2002.

58. E. Mellet, N. Tzourio, F. Crivello et al., "Functional anatomy of spatial mental imagery generated from verbal instructions," *Journal of Neuroscience,* vol. 16, no. 20. pp.6504-6512, 1996.

59. A. Amedi, N. Raz, P. Pianka et al., "Early 'visual' cortex activation correlates with superior verbal memory performance in the blind," *Nature Neuroscience,* vol. 6, no. 7. pp.758-766, 2003.

60. J. A. Ferrero Blanco, "Imágenes Acústicas y Visión Inducida: Entrevista a don José Luis González Mora." *U.T.L.A.I.Punto Doc* . 2002.

61. A. G. De Volder, H. Toyama, Y. Kimura et al., "Auditory triggered mental imagery of shape involves visual association areas in early blind humans," *Neuroimage,* vol. 14, no. 1. pp.129-139, 2001.

62. M. Behrmann, G. Winocur, and M. Moscovitch, "Dissociation Between Mental-Imagery and Object Recognition in A Brain-Damaged Patient," *Nature,* vol. 359, no. 6396. pp.636-637, 1992.

63. C. Büchel, C. Price, R. S. J. Frackowiak et al., "Different activation patterns in the visual cortex of late and congenitally blind subjects," *Brain,* vol. 121. pp.409-419, 1998.

64. B. L. Meijer, "Stereoscopic Vision for the Blind: Binocular vision support for The vOICe auditory display." *http://www.seeingwithsound.com/binocular.htm* . 2011. 4-3-2011.

65. Jie X, W. Xiaochi, and F. Zhigang, "Research and Implementation of Blind Sidewalk Detection in Portable ETA System." *International Forum on Information Technology and Applications* , pp. 431-434. 2010.

66. J. Xu and Z. Fang, "AudioMan: Design and Implementation of Electronic Travel Aid." *Journal of Image and Graphics* vol. 12 no. 7, pp. 1249-1253. 2007.

67. C. Capelle and C. Trullemans, "A real-time experimental prototype for enhancement of vision rehabiliation using auditory subsitution." *IEEE Trans.On Biomedical Engineering* vol. 45 no. 10, pp. 1279-1293. 1998.

68. G. Sainarayanan, R. Nagarajan, and S. Yaacob, "Fuzzy image processing scheme for autonomous navigation of human blind." *Applied Soft Computing* vol. 7 no. 1, pp. 257-264. 2007.

69. P. B. L. Meijer, "An Experimental System for Auditory Image Representations," *IEEE Transactions on Biomedical Engineering,* vol. 39, no. 2. pp.112-121, 1992.

70. B. L. Meijer, "Seeing with Sounds: is it Vision?" vol. Invited presentation at VSPA 2001 Conf. on Consciousness, Amsterdam. 2001.

71. P. Meijer, "Seeing with Sound for the Blind. Is it Vision? Can it be?" vol. Invited presentation at Tucson 2002, Tucson, Arizona. 2002.

72. O. Gómez, J. A. González, and E. F. Morales, "Image Segmentation Using Automatic Seeded Region Growing and Instance-Based Learning." *Lecture Notes in Computer Science, Progress in Pattern Recognition, Image Analysis and Applications* vol. 4756, pp. 192-201. 2007. Berlin Heidelberg, L. Rueda, D. Mery, and J. Kittler (Eds.).

73. G. Balakrishnan, G. Sainarayanan, R. Nagarajan et al., "Fuzzy matching scheme for stereo vision based electronic travel aid," *Tencon 2005 - 2005 IEEE Region 10 Conference, Vols 1-5*. pp.1142-1145, 2006.

74. G. Balakrishnan, G. Sainarayanan, R. Nagarajan et al., "Stereo Image to Stereo Sound Methods for Vision Based ETA." *1st International Conference on Computers, Communications and Signal Processing with Special Track on Biomedical Engineering, CCSP 2005, Kuala Lumpur* , pp. 193-196. 2005.

75. Y. Kawai and F. Tomita, "A Support System for Visually Impaired Persons Using Acoustic Interface - Recognition of 3-D Spatial Information." *HCI International* vol. 1, pp. 203-207. 2001.

76. Y. Kawai and F. Tomita, "A Visual Support System for Visually Impaired Persons Using Acoustic Interface." *IAPR Workshop on Machine Vision Applications (MVA 2000)* , pp. 379-382. 2000.

77. Y. Kawai and F. Tomita, "A Support System for Visually Impaired Persons Using Three-Dimensional Virtual Sound." *Int.Conf.Computers Helping People with Special Needs (ICCHP 2000)* , pp. 327-334. 2000.

78. D. J. Lee, J. D. Anderson, and J. K. Archibald, "Hardware Implementation of a Spline-Based Genetic Algorithm for Embedded Stereo Vision Sensor Providing Real-Time Visual Guidance to the Visually Impaired." *EURASIP Journal on Advances in Signal Processing* vol. 2008 no. Jan., pp. 1-10. 2008. Hindawi Publishing Corp. New York, NY, United States.

79. S. Meers and K. Ward, "A Substitute Vision System for Providing 3D Perception and GPS Navigation via Electro-Tactile Stimulation." *Proceedings of the International Conference on Sensing Technology* . 2005.

80. M. Khan, M. Fahad, and Siddique-e-Rabbani, "Ultrasound mobility aid for the blind using frequency modulated nerve stimulation," *2010 6th International Conference on Electrical & Computer Engineering (ICECE 2010)*. pp.171-174, 2010.

81. P. Baranski, M. Polanczyk, and P. Strumillo, "A remote guidance system for the blind," *2010 12th IEEE International Conference on e-Health Networking, Applications and Services (Healthcom 2010)*. pp.386-390, 2010.

82. A. Smailagic and R. Martin, "Metronaut: a wearable computer with sensing and global communication capabilities," *Digest of Papers.First International Symposium on Wearable Computers (Cat.No.97TB100199)|Digest of Papers.First International Symposium on Wearable Computers (Cat.No.97TB100199)*. pp.10, 1997.

83. S. Suntharajah, "Smart cane to help the blind 'see'." http://thestar.com.my/lifestyle/story.asp?file=/2005/5/29/features/20050529110045&sec=features . 2005. 12-3-2011.

84. F. Diez de Miguel, "Bastón vibrador para invidentes." http://www.inventoseinventores.com/index.php?grupo=detalles_inventor&id=312 . 2003.

85. H. B. Ceipidor, G. Azzalin, and M. Contenti, "A RFID System to Help Visually Impaired People in Mobility," *EU RFID Forum 2007, 13-14 March 2007, Brussels, Belgium.(Unpublished)*, 2007.

86. V. Kulyukin, C. Gharpure, J. Nicholson et al., "RFID in robot-assisted indoor navigation for the visually impaired," *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (IEEE Cat.No.04CH37566)*. pp.1979-1984, 2004.

87. J. Na, "The Blind Interactive Guide System Using RFID-Based Indoor Positioning System." *Computers Helping People with Special Needs, Lecture Notes in Computer Science,* vol. 4061, pp. 1298-1305. 2006.

88. T. E. Piotrowski, "RFID Navigation System." no. 1 313 079. 2003.

89. M. Petrella, L. Rainville, and D. Spiller, "Remote Infrared Audible Signage Pilot Program: Evaluation Report." vol. FTA-MA-26-7117-2009.01. 2009.

90. J. R. Marston and R. G. Golledge, "The Hidden Demand for Participation in Activities and Travel by Persons who are Visually Impaired." *Journal of Visual Impairment & Blindness* vol. Aug., pp. 475-488. 2003.

91. R. G. Golledge, J. R. Marston, and C. M. Costanzo, "Assistive Devices and Services for the Disabled: Auditory Signage and The Accessible City for Blind or Vision Impaired Travelers." vol. Report for MOU 276, ISSN 1055-1417. 1998.

92. M. Uslan, S. Malone, and W. De l'Aune, "Teaching route travel to multiply handicapped blind adults: an auditory approach." *Journal of Visual Impairment & Blindness* vol. 77, pp. 18-20. 1983.

93. M. Uslan, L. Russel, and C. Weiner, "A 'musical pathway' for spatially disoriented blind residents of a skilled nursing facility." *Journal of Visual Impairment & Blindness* vol. 82, pp. 21-24. 1988.

94. G. E. Lancioni, D. Oliva, and S. Bracalente, "An acoustic orientation system to promote independent indoor travel in blind persons with severe mental retardation." *Perceptual and Motor Skills* vol. 80 no. 3 Pt 1, pp. 747-754. 1995.

95. G. E. Lancioni, D. Oliva, and F. Gnocchini, "A visual orientation system for promoting indoor travel in persons with profound developmental disabilities and visual impairment." *Perceptual and Motor Skills* vol. 83 no. 2, pp. 619-626. 1996.

96. M. Sáenz and J. Sánchez, "Indoor Position and Orientation for the Blind." *HCI Part III, HCII 2009, LNCS* vol. 5616, pp. 236-245. 2009.

97. S. Mau, N. Melchior, M. Makatchev et al., "BlindAid: An Electronic Travel Aid for the Blind." vol. CMU-RI-TR-07-39, Robotics Institute, Carnegie Mellon University, May., pp. 1-26. 2008.

98. S. Willis and S. Helal, "A Passive RFID Information Grid for Location and Proximity Sensing for the Blind User." vol. University of Florida Technical Report number TR04-009. 2009.

99. D. A. Ross, A. Lightman, and V. L. Henderson, "Cyber Crumbs: An Indoor Orientation and Wayfinding Infrastructure." *RESNA 28th International Annual Conference 2005: Atlanta, Georgia,* pp. 1-6. 2005.

100. G. E. Lancioni, M. F. O'Reilly, N. N. Singh et al., "Orientation Technology for Indoor Travel by Persons with Multiple Disabilities." *Cognition Process* vol. 10 no. Suppl 2, p.S244-S246. 2009.

101. J. A. Hesch and S. I. Roumeliotis, "An indoor localization aid for the visually impaired," *Proceedings of the 2007 IEEE International Conference on Robotics and Automation, Vols 1-10*. pp.3545-3551, 2007.

102. J. Sánchez and E. Maureira, "Subway Mobility Assistance Tools for Blind Users." *ERCIM UI4ALL Ws 2006, LNCS 4397* , pp. 386-404. 2007.

103. D. Parkes, "NOMAD - An audio tactile tool for the acquisition, use and management of spatially distributed information by partially sighted and blind people." A. Dodds and A. Tatham, eds. *Proceedings Of The Second International Conference On Maps And Graphics For The Visually Disabled, King's College, London* , pp. 24-29. 1988. Nottingham, Uk: Nottingham University.

104. R. Dan Jacobson, "Talking Tactile Maps and Environmental Audio Beacons: an Orientation and Mobility Development Tool for Visually impaired People." *Proceedings of the ICA Commission on maps and graphics for blind and visually impaired people, 21-25 October* . 1996.

105. R. G. Golledge, J. M. Loomis, R. L. Klatzky et al., "Designing A Personal Guidance-System to Aid Navigation Without Sight - Progress on the Gis Component," *International Journal of Geographical Information Systems,* vol. 5, no. 4. pp.373-395, 1991.

106. J. M. Loomis, R. G. Golledge, and K. L. Klatzky, "Personal guidance system for the visually impaired using GPS, GIS, and VR technologies." *Proceedings of the First Annual International Conference, Virtual Reality and Persons with Disabilities* vol. June 17-18, pp. 71-74. 1993. June 17-18.

107. E. Urband and R. Stuart, "Orientation Enhancement through Integrated Virtual Reality and Geographic Information Systems." *Proc.of the SCUN Conf.on Technology and Persons with Dissabilities* , pp. 55-62. 1992.

108. M. May and Ch. LaPierre, "Accessible Global Positioning System (GPS) and Related Orientation Technologies." *Assistive Technology for Visually Impaired and Blind People*. Marion A.Hersh and Michael A.Johnson, eds. no. 8, pp. 261-288. 2008.

109. A. H. Dodson, G. V. Moon, T. Moore et al., "Guiding blind pedestrians with a personal navigation system," *Journal of Navigation,* vol. 52, no. 3. pp.330-341, 1999.

110. VisuAide, "Strider." http://www.flora.org/lapierre/strider.htm . 1994. 2-3-2011.

111. J. M. Loomis, R. G. Golledge, and K. L. Klatzky, "Navigation System for the Blind: Auditory Display Modes and Guidance." *Presence* vol. 7 no. 2, pp. 193-203. 1998.

112. T. Strothotte, H. Petrie, V. Johnson et al., "Mobic - User Needs and Preliminary Design for A Mobility Aid for Blind and Elderly Travellers," *European Context for Assistive Technology,* vol. 1. pp.348-351, 1995.

113. H. Petrie, V. Johnson, T. Strothotte et al., "MoBIC: Designing a Travel Aid for Blind and Elderly People." *Journal of Navigation* vol. 49, pp. 45-52. 1996.

114. HumanWare, "BrailleNote GPS." http://www.humanware.com/...gps/braillenote_gps . 2002. 1-3-2011.

115. S. L. Code Factory, "Mobile Geo®: The World in your Pocket." http://www.codefactory.es/descargas/family_5/Mobile%20Geo%20User%20Guide.htm . 2010. Code Factory, S.L. 1-3-2011.

116. HumanWare, "Trekker talking GPS." http://www.humanware.com/en-europe/products/blindness/talking_gps/trekker/_details/id_88/trekker.html . 2003. 1-3-2011.

117. The Loadstone GPS team, "What is Loadstone-GPS worth to you?" http://www.loadstone-gps.com/about/ . 2006. 1-3-2011.

118. K. Tsukada and M. Yasumura, "ActiveBelt: belt-type wearable tactile display for directional navigation," *UbiComp 2004: Ubiquitous Computing.6th International Conference, Proceedings (Lecture Notes in Comput.Sci.Vol.3205)*. pp.384-399, 2004.

119. Wayfinder Systems AB., "Wayfinder Access Refund Programme." http://access.wayfinder.com/ . 2010. 1-3-2011.

120. V. Ari and K. Sami, "NOPPA Navigation and Guidance System for the Blind." . 2006. VTT Industrial Systems.

121. P. E. Lanigan, A. M. Paulos, A. W. Williams et al., "Trinetra: Assistive Technologies for Grocery Shopping for the Blind." *International IEEE-BAIS Symposium on Research onAssistive Technologies (RAT '07).Dayton, OH.* vol. April. 2007. 1-3-2011.

122. N. Márkus, A. Arató, Z. Juhász et al., "MOST-NNG: An Accessible GPS Navigation Application Integrated into the MObile Slate Talker (MOST) for the Blind." *ICCHP II, LNCS 6180* vol. II, pp. 247-254. 2010.

123. Kapsys, "Kapsys Products." http://www.kapsys.com/modules/movie/scenes/learnuse/ . 2011. 2-3-2011.

124. S. K. Nagel, C. Carl, T. Kringe et al., "Beyond sensory substitution--learning the sixth sense," *J Neural Eng,* vol. 2, no. 4. pp.R13-R26, 2005.

125. A. Kooijman and M. Uyar, "Walking speed of visually impaired people
with two talking electronic travel systems." *Visual Impairment Research* vol. 2 no. 2, pp. 81-93. 2000.

126. T. Kapic, "Indoor Navigation for Visually Impaired,", A project realized in collaboration with NCCR-MICS., 2003.

127. S. Treuillet, E. Royer, T. Chateau et al., "Body Mounted Vision System for Visually Impaired Outdoor and Indoor Wayfinfing Assistance." M. A. Hersh, ed. *Conference & Workshop on Assistive Technologies for People with Vision & Hearing Impairments Assistive Technology for All Ages, CVHI 2007.* pp. 1-6. 2007.

128. B. N. Walker and J. Lindsay, "Using virtual reality to prototype auditory navigation displays." *Assistive Technology Journal* vol. 17 no. 1, pp. 72-81. 2005.

129. J. M. Loomis, R. G. Golledge, and R. L. Klatzky, "GPS-based navigation systems for the visually impaired. Fundamentals of wearable computers and augmented reality." *Fundamentals of Wearable Computers and Augmented Reality*. W.Barfield and T.Caudell, eds. pp. 429-446. 2001. Mahway, NJ, US., Lawrence Erlbaum Associates.

130. B. G. Katz, Ph. Truillet, S. Thorpe et al., "NAVIG: Navigation Assisted by Artificial Vision and GNSS." *Workshop Pervasive 2010: Multimodal Location Based Techniques for Extreme Navigation, Helsinki, 17/05/2010* . 2010.

131. A. Helal, S. E. Moore, and B. Ramachandran, "Drishti: An integrated navigation system for visually impaired and disabled," *Fifth International Symposium on Wearable Computers, Proceedings*. pp.149-156, 2001.

132. N. G. Bourbakis and D. Kavraki, "An intelligent assistant for navigation of visually impaired people," *2Nd Annual IEEE International Symposium on Bioinformatics and Bioengineering, Proceedings*. pp.230-235, 2001.

3. User Requirements

Users are not always taken into account when researchers propose new designs and devices to help them. It is, hence, important to ask the potential final users and other related experts if a practice and usable final device wants to be implemented. The standard regulation ISO 13407, "Human-Centered Design Processes for Interactive Systems" [1] proposes the inclusion of the potential users since the first developing steps of any project.

Thus, before designing the prototype, requirements from potential users and experts in some fields related with the visual disability where retrieved. This chapter exposes what and how that information was obtained.

The methodology used in this research is based on the methods used by Sinha [2], Carroll [3], Flanagan [4] and Chapanis [5]. This will be explained in detail later.

3.1. Interviews

Eleven interviews were carried out during the autumn and winter 2010, at the working place of the different experts in Madrid, Santander, Lleida and Barcelona.

3.1.1. Demographics

In the development of this study, we performed 11 one-to-one interviews to different professional and personal profiles, related to blindness, rehabilitation, psychoacoustics, computer science and music.

More in detail, the experts interviewed can be classified in the following non-exclusive categorization:

- Blind people: 6
- Psychology and rehabilitation professional profile: 2
- Technical professional profile : 4
- Experts in assistive products: 5
- Experts in music: 3

These persons belong to the following organisms, enterprises or associations:

- Universidad Politécnica de Madrid (Telecommunication Eng.)
- National Organization of Spanish Blinds (ONCE), sections of Santander, Madrid, Lleida, and individual people affiliated to ONCE.
- CIDAT: Centro de Investigación, Desarrollo y Aplicación Toflotécnica, belonging to the ONCE.
- CEAPAT: Centro Estatal de Autonomía Personal y Ayudas Técnicas.
- Universidad Complutense de Madrid (Medicine U.)
- IBM: International Business Machines.
- UTLAI: Usuarios de Tiflotecnología para el Libre Acceso a la Información.

These 11 interviews were carried out in the working centers between September and December 2010, with durations between 30 min and 130 min.

3.1.2. Structure of Interviews

The interviews were performed by means of 4 open questions and the later conversation about some specific aspects related to the given answers. Likewise, the free-listing method [2] was used in each question, giving freedom to the interviewed to list as many statements as he/she wanted.

The first question was prepared to obtain design criteria for the new device, as proposed in [3]. In this work, the Claims Analysis is described as fair technique to retrieve the problems in present and hypothetical scenarios in the users' life.

The first and second questions implement the Critical Incident Technique (CIT) [4;5], regarding daily life problems and limitations or critics to existing technical aids, which should help to a better design of the proposed device.

Before the interview, an explanation of the proposal, regarding functionality and design guidelines were sent by e-mail, in order to prevent them of the focus interest of the interview.

These 4 questions were the following:

1. Problems to solve in the blind's daily life, regarding orientation and mobility (question asked only to blind interviewed people and experts in this field).
2. Known systems or devices related to these problems, and critics to them.
3. Proposals and advices about the proposed system (at user or technical level)
4. Another contacts proposed by experts or users to be interviewed.

The last question will not be treated in this chapter, because of obvious reasons.

3.2. Results

We will follow the same order than that of the questions to present the answer and other relevant information extracted from these interviews.

3.2.1. Problems to be Solved in the Blind's Daily Life

The detected problems can be joined in three sets: (i) relative to orientation in public spaces, (ii) difficulty in the access to visual information and (iii) the avoidance of obstacles during travels.

Before analyzing in detail these sets of problems, it is important to remark that there are many standard and commercial techniques and APs designed to solve some of these problems, such as the guide-dog, the white cane or panels written in Braille. However, blind people keep detecting important lacks to solve regarding mobility and orientation problems. Most of interviewed people remarked the danger that supposes, in general terms, the mobility and displacement through public spaces, and more specifically:

- Subway stations: up step alerting the beginning of down steps in the subway entrance disappeared. This fact increments the danger when entering in the subway network.

They are, nowadays, big and open spaces, where it is difficult to find references. Moreover, noises and echoes invoke the sensation of being "in the middle of nowhere". Other public spaces, such as stations, airports, museums, shops, etc. present the same problems.

- Mobility and orientation are seen as challenges for personal autonomy and safe travel.
- Some information about the surroundings would help in displacements. That is the reason why some blind people need a company, which is still a problem to be solved.
- These problems are especially important for elderly blind people.

Every blind people, as well as experts in contact with this collective, found important challenges regarding displacements in public environments. As it can be extracted from the previous list, these problems are determined, mostly, by a lack of information in big environments and by the echo (and that fact avoids the echolocation). Such environments are not perceived as accessible or prepared for visually impaired people.

Another important set of problems is related to visual information. In this group we find the following complains:

- Written information in panels is not always accessible (street names, advices, etc.).
- Bus numbers and stops are neither accessible for visually impaired people.
- Queue ticket machines and the turn numbers are not provided acoustically nor in Braille.
- Cash machines and other automatic vending machines are rarely accessible.
- Street crossing without semaphores.
- Finding free seats in a bar or a restaurant.
- Finding semaphores and distinguishing them from lampposts.

In this group we could underline that most of the problems come from the insufficient number of panels and public information or other printed information in Braille. In another order of problems, traffic information is revealed as crucial, because the consequences of taking the incorrect choice regarding this matter, and also because not always this information is accessible. Finally, some leisure problems as finding free seats in bars was also remarked during the interviews.

The last and more detailed group of problems detected during the interviews is related with obstacle avoidance and mobility. In this group, most of interviewed people agree in classifying obstacles in terms of height, being the most dangerous ones those at the height of the head. It was assumed that the white cane is always used, even if secondary APs are proposed. This assumption makes lower obstacles less dangerous since they can be detected by the cane.

However, some lower obstacles as bollards are, sometimes, hardly perceived by the cane or the dog-guide, and provoke most of little accidents.

Other examples provided by blind people are:

- Middle height motorbikes' or cars' mirrors.
- Containers.
- Mail and telephone boxes.

- Scaffolds.

In general terms, things that overpass the wall line are perceived as dangerous. Although static obstacles can be remembered, new obstacles use to provoke many accidents.

Regarding the risk, not every obstacle is perceived as equally risky. Down steps are much more risky than up steps, for example.

Finally, interviewed blinds remember the high risk represented by electric cars, the more and the more common, because they are quiet and difficult to be detected by hearing.

3.2.2. Known APs and Critics

We will only take into account the critics that we find relevant for our study. Moreover, the systems proposed during the interviews have been already incorporated to the State of the Art of this study and, hence, will not be treated in this chapter.

There are some positive comments about some APs that are currently in use by the collective of blind people, because of the following reasons:

- Simplicity. This characteristic warranties the maximum usability for that technology. The paradigmatic example is the white cane.
- Affordable price.
- Capability to transmit useful information. Some interviewed put as example the tactile vision designed in the Complutense University of Madrid.

However, some inconvenient points were also found, regarding the existing and known APs, some inversely related to the advantages already presented:

- Price. APs systems use to be expensive, from 600 to thousands euro. Moreover, utility for final users is not always perceived as proportional to the price. Thus, the ONCE is not disposed to finance these devices.
- Weight. Some known systems were implemented over a heavy back bag, hard to be carried (as it is the case of the Acoustic Virtual Space –EAV- of the Astrophysics Institute of Canarias).
- Usability. Some proposed systems had to be managed with the hand, and this fact combined with the cane, the user could not use the hands no more. This is perceived as very negative.
- Complexity. "Users got crazy with the ultracane", because of the alarm system with beeps.
- Long trainings. This was also the case of the EAV, or the vOICe (see chapter 2 for descriptions about both systems).
- Who takes care about the sustenance of the system?
- The potential users for technical APs are few, and there are no scale economies in this field.

We can underline the low knowledge that users have about proposed APs. In the State of the Art chapter there are around 75 APs explained or cited which were designed for mobility. With the exception of a rehabilitation technician, the rest of interviewed people could hardly cite around 5 or 6 of them. Thus, one of the main problems found with the proposed APs is the ignorance of the public of their existence. This barrier seems to be broken more easily in the same country, such as happened with the EAV.

Preferences seem also to be clear enough. The average user of APs in the blind collective is not an enthusiast of technology; hence, these APs cannot overpass a certain level of complexity, neither in the training nor in the daily use. Moreover, the price is one of the most important constraints for their implantation, standardization and use, since the ONCE, in Spain, has some restrictive criteria regarding this factor.

Finally, experts underline the lack of usability and portability as a crucial problem of existing APs. Two are the main reason: On the one hand, because of the lack of comfort. On the other one, because no blind person wants to walk in the street with showy hardware mounted over his/her person.

3.2.3. Proposals and Advices about the Proposed AP

The answers given to this question must be divided in two main groups: those regarding usability, and those more techniques.

Related to the previous critics, and at a user level, we found a huge number of advices and alerts. We will divide, for presentation purposes, the answers in subsections.

Usability:

- Possibility to integrate the system in a cane, in order to use the hand to orientate the device to the representative area.
- Water resistant.
- Easy to operate.
- Handy orientation angle, not only with the head.
- Adaptable, portable, easy to handle.
- Easy to be operated by elderly people with perceptive problems. For elderly people it must be very simple in order to avoid rejects.
- Taking into account people who do not hear very well.
- Possibility of different profiles in the design. Implementation of some of them in the final prototype. Adaptable to different perceptive characteristics.
- Moderated complexity.
- “Usability is borderline” (Nielsen).
- Look forward less complexity and more functionality.
- Take into account the final population:
    - Total blinds?
    - Partial blinds?
    - With intellectual disabilities?
- The final group should be as wide as possible, even if the final AP will only be used by a small fraction of it.
- Not everybody uses in the technology in the proper way.
- The design based on glasses should be adaptable to anyone, given the blind rejection to ostentatious apparatus.

- The system must be portable. The hardware format is very important. It should not be a back, heavy or showy.
- The tiredness it may cause must be taken into account.
- The functioning during the training may be different than that of the normal use. The interaction model may change once the training period is over. It is important to pay attention to the capabilities required for the training.
- Carefulness with the expectatives created, because they could yield to frustration.
- A child should be able to know the form of thing that he/she cannot touch.

As it can be seen in the answers, comfort in the use is perceived as a capital aspect of the final implementation. In another direction, experts alert about deafblindness and elderly people. These collectives present special cognitive capabilities and necessities, which should be taken into account if we aim their integration.

Likewise the training is perceived as an important challenge, advising to keep it simple and natural in order not to exclude anyone.

Information channels:

- Reverberation is unpleasant, disturbing the perception.
- The ear cannot be supplanted. Sounds should not interfere because it blocks the echolocation. Moreover, it can diminish the entrance of ambient sounds.
- Many sounds could be unaffordable for the brain.
- “May the system recognize bulks such as semaphores?” Form recognition with different timbres and frequencies. “Table should be a single sound”. This should be implemented with specific forms and objects.
- Alert with relevant and imminent obstacles: holes, semaphores, train… Other things are not relevant, “just bulks”.
- An AP must bring extra information, without eliminating other information.
- Possibility of recognizing colors, even if it must be only in the center of the image.
- Vibrations should be used because they interact less with the ambient noise. Different vibrations may be used: impulses, continuous… Stereo vibration may be use to indicate where is the obstacle. Find sensible areas of the body in which place the actuators and find the accuracy. The main advantage is that it would be valid for deafblind too. However, regarding to the amount of information “better 40 than 4” if sounds can be used in a proper way.
- “We don’t listen to everything we hear”. Attention plays an important role.
- The system may adapt to the ambient noise, and be interactive with the surrounding.
- The transmitted information must be selected. It is mandatory to build a hierarchy of information to be shown. For example, “walking people are not so important”.
- There are differences between canes and guide dogs, regarding obstacles avoidance. Those who use canes could be more interested in the proposed AP.
- Separate orientation from mobility, even if a final integration should be very positive.
- Combination of local and global information.
- Technology must be adaptable to circumstances. Do not saturate the channel.

- Decide de final channel to be used:
    - Acoustic
    - Vibration
    - Visual if some residual sight is available?
- To reduce de sample of the environment, it is positive to mount the device in the head. If the goal is not substituting the cane, the sampling can be yield only in middle height.
- In invasive surroundings it is important not to saturate the user with too much information.
- The system must warranty that no critical omissions are allowed.
- It might help to define dangerous situations and recognizing them.
- In the street there is too much noise; vibration should be useful in these situations. Bone transmission might also work in this environment.
- Maybe a micro loudspeaker hanging from the glasses are a solution in order to respect the hearing system.

Firstly, there is a complete unanimity among the blinds regarding the necessity of respecting the ear and the hearing system, given that it is the most important channel for incoming information. Thus, the reject to earphones is, likewise, unanimous.

However, there is a positive reception to the use of bone transmission, as well as other channels.

The use of vibrations is perceived as very advantageous, regarding some special collectives or environments. However, it is also perceived the intrinsic limitations of vibrations regarding the hearing system.

Another important advice given by experts and users is the necessity of selecting correctly the information to be transmitted to the user, as well as how it is transmitted. The most important challenge is related with the difficultness of interpretation, as well with the possibility of channel saturation and, hence, the user.

Interviewed people have shown to be favorable to a certain capacity of the system to recognize forms and objects, to differentiate some obstacles.

<u>Training and tests:</u>

Some advices are related to the training and tests bench. More in detail:

- Tests the worst case, not only in the labs.
- Tests with children (from 8 years-old), because they are who finally establish new technologies.
- Tests the "vision" generated by the system by means of EEG or other brain imaging techniques.
- Define the training period and estimate its length. This datum should be evaluated by experts and rehabilitation technicians. Known and unknown surroundings should be defined, to perform tests with controlled variables, control groups, etc.

We can easily see the constant concern regarding the training or learning period, given that, in some implementations, it has shown to be the bottleneck of the whole project.

Sonification:

Sonification is the process to convert some data to sounds. In our case, the input data are images, hence, sonification is the conversion of images to sounds representing such images. Technicians and some blind people answered to this question in this sense. The explained proposal aimed to convert images with depth information into sounds.

- Two clearly different timbres. Maybe some simple combinations of a purr with a flute (for example), completing the code later.
- Buzzing is easily perceived.
- Frequency has much information, but we should pay attention to timbres.
- Combination of several sounds/frequencies may bring a lot of information.
- Musical notes might not be useful.
- The mobile phones sounds and melodies are good clues, because of their variety, as well as musical instruments.
- A figure should have assigned an involving sound.
- Soft sounds are better, such as human voices, natural sounds, water sounds…
- The set of sounds should not bother. It is important not to use unpleasant or invasive sounds. They should be uncorrelated with the external sources.
- Orchestra instruments are incorporated because of our culture.
- Varying the working cycle of a square wave we can generate many harmonics.

First of all, we see some contradictions among the different proposals received by the experts; for example, regarding the suitability of using or not musical notes. An emphasis has been also given to the possibility of take advantage of patterns culturally assumed, such as human voices, orchestras, etc, given their familiarity. Last, all interviewed people seem to agree in the wide possibilities of the frequency as information carrier, as well as the different timbres (by the way, closely related with frequency).

An important advice has been repeated in several interviews:

- Unpleasant noise and/or vibration alarms to advice of imminent dangers.
- Boolean alarm to avoid crashes.

The possibility of combining sounds and vibrations had already appeared before. The necessity of unpleasant noise to alert of imminent crashes, combined with vibration alarm seems to be unquestionable, since the priority of this system is to avoid accidents, and not allowing to “see” small details, mostly irrelevant for mobility.

3.2.4. Other Issues Extracted from Interviews

In the following list, we will present other proposals or comments, classified regarding the parameters and dimensions.

Distance:

- From further to closer objects, the sound might change (with intermittent beeps, for example).
- Further than 3 meters the information is irrelevant.
- The system might describe the bigger obstacle, or the closer, or the most dangerous.
- The depth might be codified in the volume, even if it gives an idea of the drawing and not of the real distance.
- It might add timbres' characteristics. The range should be gradable.

Horizontal axis:

- This axis is related to spatial perception. It is important to use the stereo effects.
- The wide of an object is hard to be imagined. It could be codified with two harmonic sounds.
- The wide of an object could be described with two sounds for the sides of the object.
- Spatiality based on binaural sounds.
- There is a dependency: frequency (f)<1 KHz, the delay is the most relevant. If f>1 KHz, the intensity is prevalent. Like this, sources are placed in the horizontal axis.

Vertical axis:

- Frequency for the vertical axis.
- Height: children frequencies are higher, adults' and fat people's frequencies are lower: "Lo agudo es pequeño".
- The low sounds might be in the bottom, the high ones at the top of the image.
- In the vertical axis, the head movements are used to place the sound sources.

Temporal variations:

- It could be tested continuous versus intermittent.
- Timbre is useful. A "vibrato" of a musical note could be related with the movement, linking its amplitude and velocity of the displacement.
- When the image is standstill, it should provoke no sounds (implement a differential detector).
- It could be a beep varying its pattern in relation to the depth.

Echolocation:

- It is proposed to use a reference melody and to use the pixels (the depth) as an echo source in a room. Configure the pixels as sources.
- With delays and intensities, objects can be localizable sources in the space.
- Modifying the natural characteristics with the displacement of acoustic sources.
- Taking into account the transmission of the head.

Limits of the system:

- Less than 17 different sounds in any case. We must distinguish between mobility (less than 6 sounds) and vision (less than 17).
- Maybe, it is easier to start with samples, pre-recorded chords; it might be more pleasant and easier.
- The system must have ON/OFF automatic capability: fading in three dimensions.

Splitting the contributions in different dimensions helps us to analyze in detail this information.

Regarding the depth, which is the basic parameter of the system, we received advices proposing the use of the volume, as well as timbres' characteristics, given that, by its own means, the volume cannot give a real measure of the distance, but just a relative one. If the system includes an Automatic Gain Control (AGC) to be able to change in relation with the ambient noise, the volume, even if it contributes with some information, cannot be used for realistic depth measurements, since it is modulated by the AGC block. Thus, it might be complemented with the timbre. However, we cannot renounce to use the volume to express distance, because its relation to real world's distances is evident and, hence, natural. Another advice in this sense is the uselessness of information of objects further than 3 meters. As simplicity is a priority of the design, we must discard most of the pixels in most of the situations; all of them with intensity lower than a certain threshold. Moreover, it will rebound positively on the efficiency of the system, because there is less information to be processed.

Regarding the horizontal axis, there is unanimity about the use of stereo and binaural sounds to express horizontal position of objects. However, there are some difficulties such as perceiving the wide of the objects, representing, or not, only the lateral bounds or the whole set of pixels of such object. Interviewers have recalled the psychoacoustic properties of the hearing system, regarding binaurality.

The vertical axis presents specific problems. On the one hand, lower accuracy when localizing sound sources, depending of the model of head and ear transmission, as well as the relative position of the head in relation with the object. Thus, some arbitrary proposals, as using the frequency to represent the height, were found. However, in this proposal we found contradictory possibilities. On the other hand, we found solutions proposing a temporal variation of generated sounds as vehicle of information. Such variations can be intermittent beeps or "vibrato" modulations codifying the relative movements of objects. Moreover, there has been proposed that only moving objects or scenes should be codified as sounds, producing static images silence. This implies a temporal differential filter and optical flow analysis, which may overload the computing tasks of the system. Another problem with the temporal variations of the sonification is that the system does not correlate an image to a static set of sounds, achieving authentic real time performance, but codifications of information during seconds, depending of the sound represented not only of the current image, but also of the previous ones. An intrinsic delay appears.

Remembering that echolocation is an important aspect for blind people's orientation, there have been proposed some methods dealing with psychoacoustics properties of the head and hearing systems, specifically, those related with the so called head-related transfer function

(HRTF). This possibility allows localizing sound sources in a 2D space, with some limitations [6]. The stereo sounds, indeed, already take into account this psychoacoustic property. But regarding the vertical axis we should discuss if it is worth enough to use an arbitrary representation of the information in this axis, providing more accurate information without a heavy training period.

In general terms, we can find global advices regarding some limits of the system: no more than 6 sounds for mobility (i.e. avoiding obstacles) and no more than 17 for a “vision” system. These data are related with the functional accuracy of the system. Another general advice recommends using samples instead of synthetically generated sounds.

3.3. Discussion

The set of interviews has provided a deeper analysis of the daily life and problems of the blinds, pointing to some critical aspects, which should be taken into account during this Ph.D. Thesis work. The variety of professional profiles interviewed generates a global landscape about the issue. An important problem of this process of gathering information is the relatively reduced sample, and the possible bias of the conclusions extracted from it will incarnate the rest of the design process.

Regarding to the matter of the interviews, we have found some incongruence and diverging lines when focusing on some specific aspects, where more subjective positions arise, instead of technical ones. This is consistent with the design process, since not every step is technically justified, and some arbitrary decisions should deal with these subjective aspects.

The decisions will be based, in these cases, on some other similar APs designed, if this reference is consistent with our requirements. If no incongruence is notified, we will follow, as far as possible, the proposals given by the experts.

On the contrary, some other points of view seem to be widely spread:

- Simplicity. The lack of simplicity of a system is seen as one of the main constraints to use a new device.
- Price: Prices are perceived as very high, obstructing the commercialization.
- Comfort: Experts state that a comfortable device is mandatory in mobility, since it is used long time each day.
- Training: The training is also perceived as a barrier to access to a service, overall for elderly people or other somehow limited potential users.

Much detailed information was also extracted regarding different acoustic representations of objects or about the hardware itself, among others, which will be recovered in the following chapters.

Taking these results and guidelines, we will build a proposal in the following chapter.

References

[1] ISO: International Standard ISO 9999. Assistive products for persons with disability — Classification and terminology. (2007)

[2] R. Sinha, Beyond cardsorting: Free-listing methods to explore user categorizations, Boxes and Arrows, 2003.

[3] J.M. Carroll, M.B. Rosson, Getting around the task-artifact cycle: How to make claims and design by scenario, ACM Transactions on Informational Systems. 10, pp. 181-212, 1992

[4] J.C. Flanagan, The critical incident technique, Psych.Bull. 51, pp. 327-358, 1954.

[5] A. Chapanis, Research techniques in human engineering, John Hopkins Press, Baltimore, Maryland; Ramsey (1977), 1959.

[6] M. Pec, M. Bujacz, P. Strumillo, A. Materka.: Individual HRTF Measurements for Accurate Obstacle Sonification in an Electronic Travel Aid for The Blind. International Conference on Signals and Electronic Systems (ICSES 2008). pp. 235-238, 2008.

4. Proposal

In the frame of this Ph.D. thesis, a new system to help blind people in their displacements is proposed. Likewise, some problems have been detected in both the State of the Art and the Users' Requirements chapters, by mean of a review of scientific literature and interviews to blinds and experts in blindness and technological aids.

These problematic aspects and constraints, as well as some positive clues found in the literature and interviews will be taken into account to define the functional and schematic structure of the proposed AP. However, some aspects will not be treated in this chapter but in the following ones in a more specific way.

4.1. Problem Description

As extracted from the Interviews summarized in the Users' Requirements chapter, blind people still detect important problems in their daily live, regarding secure displacements. Among the detected problems, getting oriented in open and public spaces and avoid unexpected obstacles seem to be the most important ones, since they appeared in almost every interview to blind people and other experts in the visual disability field.

Several technological proposals have been described, some of them commercial, others just prototypes, to help blind people in these daily tasks, which have been reviewed in the State of the Art chapter. However, some problems have also been detected regarding these proposals, being the most important:

- Usability
- Portability and discreteness
- Price

The first aspect is studied in detail by Trefler and Hobson in [1]. In this work, some relevant aspects are highlighted regarding usability:

- High versus Low technology: this is a non-static definition. Grossly, it refers to non-electronic and electronic devices. In these terms, the white cane would be low technology, while a mobile phone falls in the high technology set. However, the more the technology devices become familiar and public, the threshold between these two groups can be moved. It is interesting, for our study, to notice that high technology uses to provoke higher rejection ratios among the public, and blind people are not an exception for that rule, as stated in the users' requirements. The practical problem is not whether a device is high of low technology, but the interface. Here, usability becomes the crucial point. It does not matter how the device is built up; the interface must be as simple as possible, as it is the case of the cane.
- Minimal versus Maximal technology: Following the first distinction, minimal technology would be, following the authors, simple devices helping to develop some task, like long-handled sponge, while maximal technologies would be devices that can

perform complex tasks, like robots. This aspect is also related with usability, independently, however, of the complexity of the task performed by the AP.

- Appliance versus Tool: The main difference of appliance and tools is that the first group does not need any special skill, capabilities or training to be used. That would be the cases of eyeglasses. Contrary, a tool needs specific skills to be useful. In our terms, the more “appliance” is an AP, the more usable it will be. However, since new and high technology is applied in modern APs, all of them partially present a “tool” facet, which should be as simple as possible. This aspect is directly related with the training process, and will be taken into account when designing the interface.

The portability has proved to be another important lack of some assistive products, sometimes because they need both hands (or at least one hand) to be used, other times because they need heavy back bags or portable PCs, showy helmets, etc. Comfort and discreteness are, somehow, related. But they are not the same. A discrete or automatic device, carried in a heavy bag will be as rejected as a very light device which needs both hands and, hence, does not allow using the cane or the guide dog. Discreteness, automatic and lightness are three crucial characteristics of any AP which aspires to be widely used.

Regarding the price, we found that, in the case of the Spanish Blinds National Association (ONCE), an important constraint to decide whether or not to finance the acquisition of APs is the price. Returning to Trefler’s and Hobson’s work, another difference in APs can be remarked, related to the price: Custom VS Commercial technology [1]. How a new device is offered to the user can follow different research and economical paths. A new device can be designed to very specific or even individual use of somebody and, hence, we can talk about custom design or fabrication. On the other hand, if a device is oriented to a bigger group of users, or even for everybody (as it is the case of the “accessible for all” paradigm), the device aspires to be commercial. The economical consequences of these two options are relevant, since scalar economics can be applied to the second case, while devices following the first way use to be very expensive (which, at its turn, makes difficult their commercialization). The same difference can be applied to the components used to implement a new device. The more commercial they are, the cheaper will be the final device, even if it is custom technology.

4.2. Baseline of Similar APs

The proposed AP is partially based on some previously implemented, published or even commercialized devices, which present some problems that we will try to overcome.

In the State of the Art, many APs were referred and described for instance the head mounted devices for mobility, among them. We will focus on this specific field to find models and ideas, which would be useful for a novel proposal.

We found a very useful idea in the Sonicguide, shown in figure 4.1.

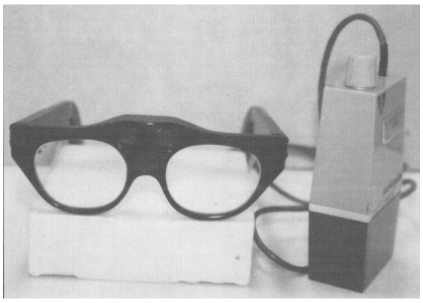


Fig. 4.1. The Sonicguide, from [2].

Other technical aids proposed in this direction were, among others:

- Sonic Pathfinder [3]
- Tyflos [4]
- Echolocation [5]
- vOICe [6]
- FIU Project [7]
- 3-D Space Perceptor [8]
- NAVI [9]
- SVETA [10, 11]
- AudioMan [12]
- CASBliP [13]
- EAV [14, 15]
- 3-D Support System [16]
- Brigham Project [17]
- The binaural sonar [18]
- The Wearable Collision Warning System [19]

Only the price of the vOICe (< $500 [20]) and an estimation of the price of the CASBliP (~1000€, “similar to a basic audiphone”) were found. Tyflos and the Wearable Collision Warning System state they are low-cost, but no price is given.

The main idea of all of them was the image or ultrasound processing to convert them into sounds, which could help the user to get oriented in the micro navigation.

The basic scheme can be found in [21], and it is shown in figure 4.2.

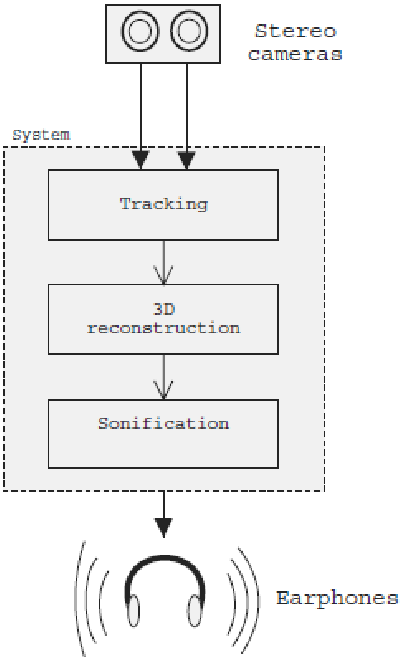


Fig. 4.2. The cross-modal electronic travel aid, from [21].

Remembering the main constraints for public utilization, such as portability and discreteness, the best designed AP in this line, regarding our guidelines, is presented in [17], and shown in figure 4.3.

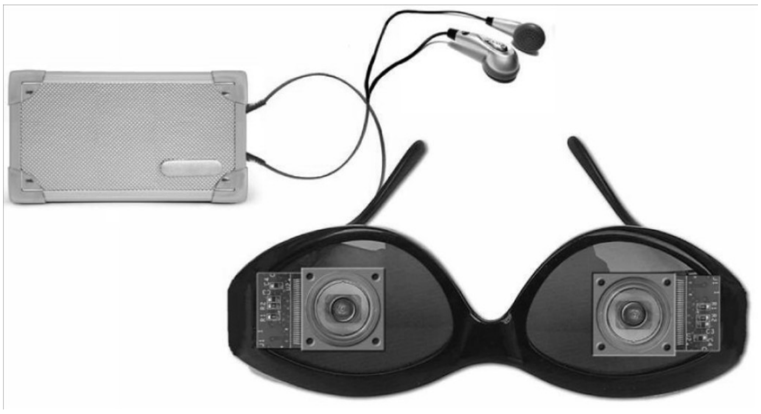


Fig. 4.3. The Real-time guide for the visually impaired, from [17].

Although this figure is an anticipation of what Lee *et al.* look for, the idea of embedding the visual sensors in some common glasses, connected to a compact processor is worthy. However, the way this information is provided to the user in this project, as well as in many other (like the SVETA, the 3D-Support system, the vOICe, the CASBliP, the EAV, and others), crashes to an unavoidable claim of the potential users: to keep ears clear. This claim will have to be taken into account in our proposal.

Given all this information, we are in disposition of proposing a functional block of an AP to help people in their mobility.

4.3. Architecture

According the main objective of this work, the architecture must try to fulfill the user´s requirements (chapter 3) as well as aspects related to ergonomic, usability, price, among others.

In figure 4.4 there are shown the functional and information flow schemes of the proposed system.

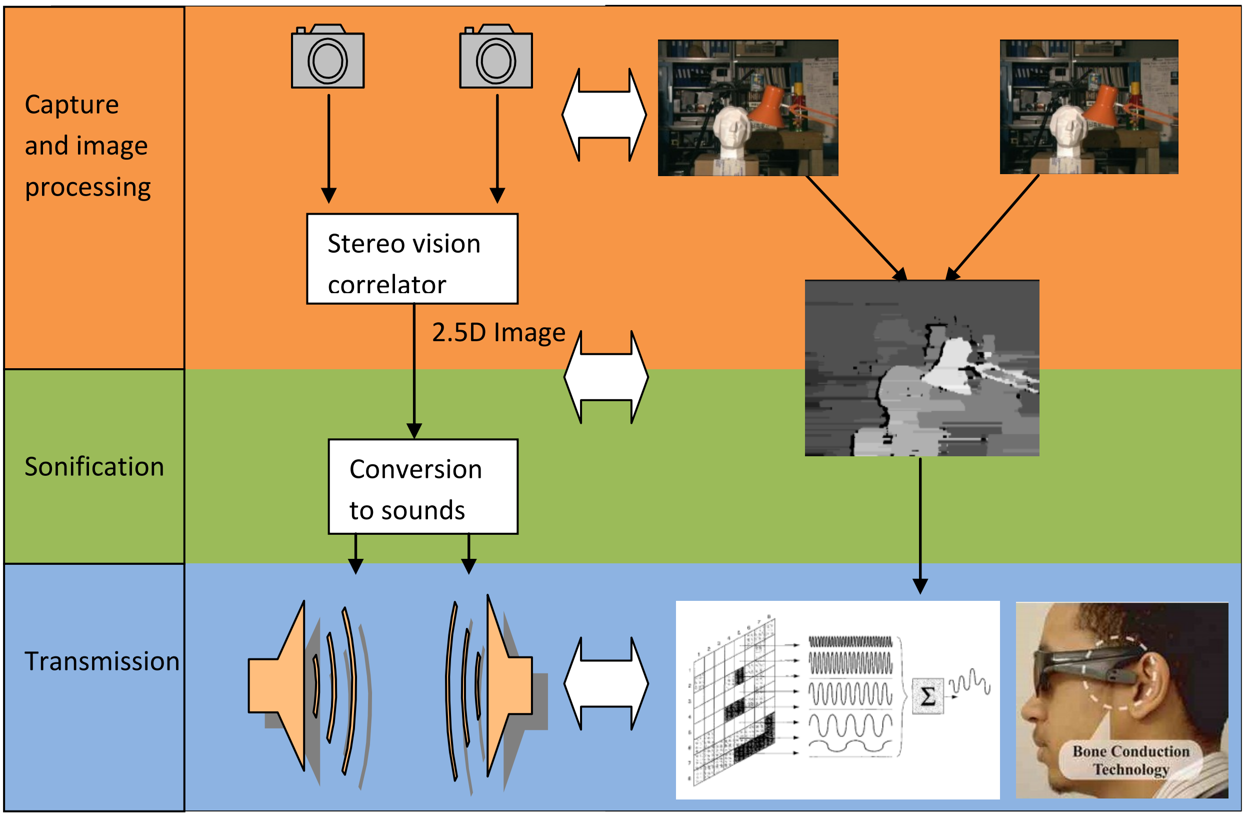


Fig. 4.4. Functional and informational schemes of the proposed AP.

The system is supposed to work in the following way:

- Two images are captured with cheap and commercial microcameras, following the stereovision paradigm (see chapter 5).
- A correlator extracts the depth map of the captured pair of images and has an image in 2.5D format as output. This image is a grayscale picture with information about the distances of each pixel to the cameras.
- A sonificator, i.e., a block that converts images to sounds, process the 2.5D image and generates an adequate pair of sounds (given that a binaural sonification is proposed).
- A final transmitter sends the acoustic information to the user, by means of a bone conduction device.

4.4. Specific Objectives

An assistive product deals with many different aspects and knowledge fields, as seen in the previous sections. Thus, several investigations and proposals must be implemented to solve different and specific problems detected in each area and field.

We must propose several partial objectives, depending of the research field. In this sense, the proposed specific objectives will be divided into image processing, sonification, transmission and training objectives.

4.4.1. Image Processing

In the following chapter (chapter 5) we will analyze the main problems around the image processing, overall regarding computational costs, complexity and speed, which are highly related.

Thus, we can already propose some objectives to be prosecuted and reached, in order to implement a portable, cheap and functional system. We can summarize them in 3 main items, proposing, as well, some absolute constraints to these variables:

- Low computational complexity: Related with the previous one, the image processing algorithm must be as simple as possible. The use of memory, given that it is usually related with the cost, should be also as small as possible. This variable will have to be, as well, minimized, and never higher than 4 times the original image, including all variables needed. For a 320×240 gray scale image, 307.2 KB.
- Real-Time constraint: An important condition of the image processing system will be to process images close to real-time conditions, i.e., at 24 frames per second.
- Accuracy > 75%: In order to avoid miss-detection of dangerous objects, nor false positives, accuracy over 75% should be achieved.

4.4.2. Sonification and Transmission

The sonification is the process by which an image is converted into sounds. This process, as it will be shown in chapter 6, presents many degrees of freedom. Thus, some approaches can be more useful for our purposes than others.

In these terms, we can establish the following specific objectives for the sonification subsystem:

- Real-time constraint: As done in the image processing section, the real-time constraint is mandatory in order to provide updated and softly-changed acoustic information to the user. Indeed, this subsystem must be synchronized with the image processing system to perform a unique task, from the user's point of view.
- Intuitive code: The sounds used to codify the image should take advantage of the natural hearing capabilities.
- Accurate descriptors: Despite the previous objective, accurate information must be provided to the user, regarding distances and spatial positions of detected objects and volumes. Accuracy should not be freely sacrificed by simplicity.

Regarding the transmission of the acoustic information, some important objectives should be reached:

- Complexity allowance: The channel to be used must have a wide enough bandwidth. This means that codes over Boolean alarms should be allowed.

- Hearing system respect: Since blind people valorize the hearing system as their most important sense, the proposed transmission channel must respect this information entrance, not overlapping the natural world sounds.
- Boolean alarm: Independently of the channel, a Boolean alarm that could be always perceived should be implemented, to avoid danger situations in any conditions or environments.

4.4.3. Global system

Important and objective parameters of the final system are the global price of the system and accuracy, which should be kept as low/high as possible, respectively. Thus, it is important to use low cost hardware in every step of the information processing chain and, hence, also in the image sensors and processors. Moreover, we must keep in mind the final application of the image processing algorithm, i.e., to detect relevant obstacles. Hence, it is not necessary to process objects below a certain size. This objective simplifies the performance of image sensors needed, since low-resolution images are enough to detect relevant objects. As stated in [4], talking about the Tyflos system, "the stereo cameras create a depth map of the environment […]. A high-to-low resolution algorithm drops the resolution of the depth map into a low resolution" image suitable for sonification.

Moreover, the processors should be commercial and powerful enough to achieve the other constraints without increasing significantly the global price.

Additionally, we can propose a minimum autonomy of the device fully working (processing images and sonifying) of around 3 hours. Few displacements in daily live exceed this time, needing permanent mobility information.

We can finally define a ratio between speed, accuracy, memory used and price that should be maximized, taking into account some absolute constraints which will be exposed later. This ratio is shown in the following expression:

$$e = \frac{speed \cdot accuracy \cdot autonomy}{price} = autonomy \cdot (delay \cdot errors \cdot price)^{-1} \tag{4.1}$$

The global "merit" of the image system "e" will be directly proportional to the speed and accuracy, and inversely proportional to the price.

Eq. 4.1 ignores the relevancy of memory since its contribution to the viability of the system is not at all relevant.

Setting a limit for the price of 500€, the merit of the worst-case system will be:

$$e = \frac{24 \cdot 0.75 \cdot 3}{500} = 108 \cdot 10^{-3} \tag{4.2}$$

4.4.4. Training

A final important section of specific objectives, we must focus on the training process. As stated in [22], acquiring mobility skills is not easy even for young students with acute senses. For older people, this process may be even impossible. Thus, the training is perceived as a bottleneck of the AP implantation process, which is also deeply related with usability.

There are skills and capability differences among potential users. These differences and possible consequences were found during the interviews and, hence, should be included in the specific objectives regarding the training process:

- Different levels/profiles: To allow every potential user to perform a training period with the proposed AP, different profiles and levels should be implemented and proposed, with an inclusion architecture, from simplest to more complex interfaces:

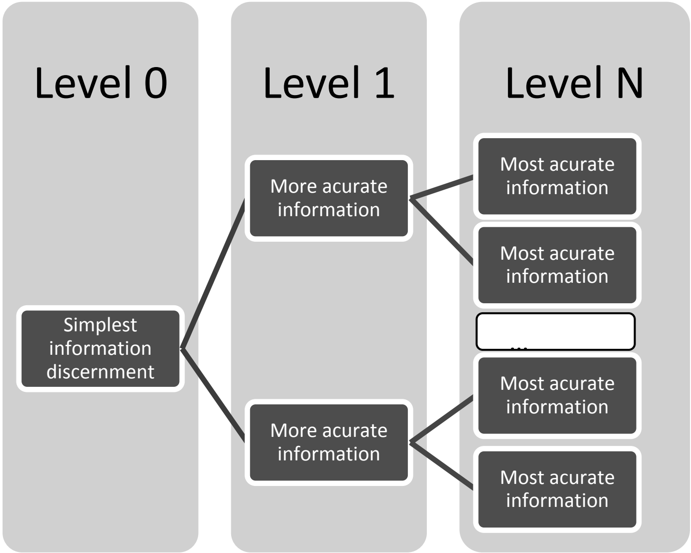


Fig. 4.5. Hierarchical profiles architecture.

As an example of the proposed nesting structure, figure 4.6 shows the proposed division, applied to gray scale levels.

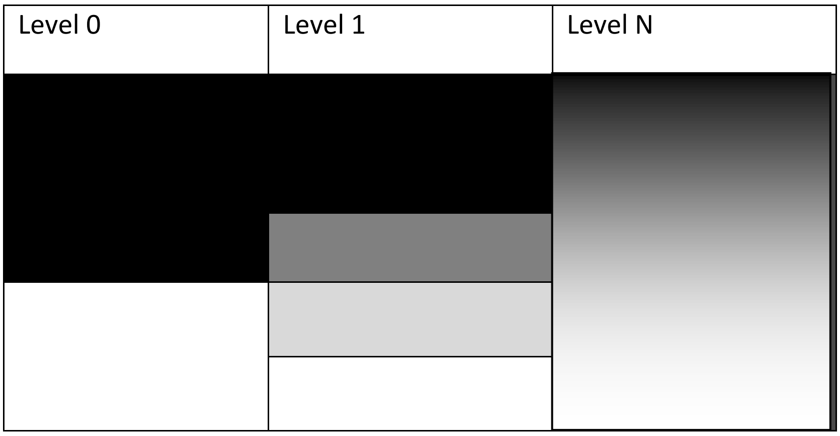


Fig. 4.6. Gray scale hierarchical classification.

- Exploitation of intuition: Combined with the previous objective, the training should take advantage of intuition, to ease the training and, hence, the use of the AP.
- Daily live based: Given that the daily surroundings are familiar for every people, including the blinds, these scenarios should be used as training environments, which could help the blind to embed the sonification process in the unconscious mechanisms of the brain.

4.5. Conclusions

After detecting some important lacks of the already reported assistive products, we are in disposition to propose a new system. In order to include the users since the beginning of the design, we performed interviews with experts of different fields related with our problem, as well as potential users. This useful information has been taken into account to design the system that will be developed in the following chapters. The proposed approach to the mobility aid tools world will try to overcome some disadvantages and problems found by the blinds in the current and commercial APs.

Moreover, we have defined some specific objectives of each branch of the system, dealing with the global price, usability and ease of use and training.

References

1. E. Trefler and D. Hobson, "Assistive Technology." C.Christiansen & C.Baum, ed. vol. Occupational therapy: Enabling function and well-being (2nd Ed.) no. 20, pp. 483-506. 1997.

2. J. A. Brabyn, "Mobility Aids for the Blind." *Engineering in Medicine and Biology Magazine* vol. 29 no. 4, pp. 36-38. 1982.

3. T. Heyes, "The domain of the sonic pathfinder and an increasing number of other things." *From http://www.sonicpathfinder.org* vol. 4.15. 2004.

4. N. G. Bourbakis and D. Kavraki, "An intelligent assistant for navigation of visually impaired people," *2Nd Annual IEEE International Symposium on Bioinformatics and Bioengineering, Proceedings*. pp.230-235, 2001.

5. T. Ifukube, T. Sasaki, and C. Peng, "A Blind Mobility Aid Modeled After Echolocation of Bats," *IEEE Transactions on Biomedical Engineering,* vol. 38, no. 5. pp.461-465, 1991.

6. P. B. L. Meijer, "An Experimental System for Auditory Image Representations," *IEEE Transactions on Biomedical Engineering,* vol. 39, no. 2. pp.112-121, 1992.

7. D. Aguerrevere, M. Choudhury, and A. Barreto, "Portable 3D sound / sonar navigation system for blind individuals." *2nd LACCEI Int.Latin Amer.Caribbean Conf.Eng.Technol.* pp. 2-4. 2004.

8. E. Milios, B. Kapralos, A. Kopinska et al., "Sonification of range information for 3-D space perception." *IEEE Transactions on Neural Systems and Rehabilitation Engineering* vol. 11 no. 4, pp. 416-421. 2003.

9. G. Sainarayanan, R. Nagarajan, and S. Yaacob, "Fuzzy image processing scheme for autonomous navigation of human blind." *Applied Soft Computing* vol. 7 no. 1, pp. 257-264. 2007.

10. G. Balakrishnan, G. Sainarayanan, R. Nagarajan et al., "Fuzzy matching scheme for stereo vision based electronic travel aid," *Tencon 2005 - 2005 IEEE Region 10 Conference, Vols 1-5*. pp.1142-1145, 2006.

11. G. Balakrishnan, G. Sainarayanan, R. Nagarajan et al., "Stereo Image to Stereo Sound Methods for Vision Based ETA." *1st International Conference on Computers, Communications and Signal Processing with Special Track on Biomedical Engineering, CCSP 2005, Kuala Lumpur* , pp. 193-196. 2005.

12. J. Xu and Z. Fang, "AudioMan: Design and Implementation of Electronic Travel Aid." *Journal of Image and Graphics* vol. 12 no. 7, pp. 1249-1253. 2007.

13. D. Castro Toledo, S. Morillas, T. Magal et al., "3D Environment Representation through Acoustic Images. Auditory Learning in Multimedia Systems." *Proceedings of Concurrent Developments in Technology-Assisted Education* , pp. 735-740. 2006.

14. O. Gómez, J. A. González, and E. F. Morales, "Image Segmentation Using Automatic Seeded Region Growing and Instance-Based Learning." *Lecture Notes in Computer Science, Progress in Pattern Recognition, Image Analysis and Applications* vol. 4756, pp. 192-201. 2007. Berlin Heidelberg, L. Rueda, D. Mery, and J. Kittler (Eds.).

15. L. F. Rodríguez Ramos and J. L. González Mora, "Creación de un espacio acústico virtual de aplicación médica en personas ciegas o deficientes visuales." vol. From: www.iac.es/proyect/eavi/documentos/EXPBEAV_25v1.DOC. 1997.

16. Y. Kawai and F. Tomita, "A Support System for Visually Impaired Persons Using Acoustic Interface - Recognition of 3-D Spatial Information." *HCI International* vol. 1, pp. 203-207. 2001.

17. D. J. Lee, J. D. Anderson, and J. K. Archibald, "Hardware Implementation of a Spline-Based Genetic Algorithm for Embedded Stereo Vision Sensor Providing Real-Time Visual Guidance to the Visually Impaired." *EURASIP Journal on Advances in Signal Processing* vol. 2008 no. Jan., pp. 1-10. 2008. Hindawi Publishing Corp. New York, NY, United States.

18. 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.

19. B. Jameson and R. Manduchi, "Watch Your Head: A Wearable Collision Warning System for the Blind," *2010 IEEE Sensors*. pp.1922-1927, 2010.

20. B. L. Meijer, "vOICe Web Page." *http://www.seeingwithsound.com/winvoice.htm* . 2013.

21. F. Fontana, A. Fusiello, M. Gobbi et al., "A Cross-Modal Electronic Travel Aid Device." *Mobile HCI 2002, Lecture Notes on Computer Science* vol. 2411, pp. 393-397. 2002.

22. E. J. Gibson, *Principles of perceptual learning and development*: New York: Appleton-Century-Crofts, 1969.

5. Image Processing

The Image Processing (IP) field refers to the art and techniques developed to process images and/or video. The set of algorithms implemented in this area allows compressing, modifying, interpreting, generating, analyzing, denoising or classifying, among other operations, widely studied in the literature.

Image processing is a very important issue to develop automatic and autonomous systems, since it allows interacting with the environment. The main problem found when implementing algorithms in this area is the big amount of information needed to be processed, producing a computational load that, in general terms, make expensive to deal with this kind of information. Thus, optimization in programs and algorithms is a capital task when a real time implementation needs to be achieved.

As we can find in any other field of engineering, optimizing has a cost that forces us to reach a trade-off between accuracy and speed. The position of this equilibrium will be set by the constraints of the specific application. No a priori directive to solve this situation is given by the research area itself. As explained in the Introduction chapter, due to the very specific application of our proposal, the speed will be the heaviest constraint, neglecting the accuracy below a certain threshold, since the final resolution is not crucial.

As explained in the Introduction chapter, the image processing is a main part of the research and development done in this thesis work. In our specific approach, a narrow branch of the image processing world will be reviewed in detail, that part regarding to 3D vision.

5.1. 3D vision Fundamentals

“3D vision” refers to the depth perception of a three dimensional scene. In figure 5.1 an image of a 3D scene is shown:

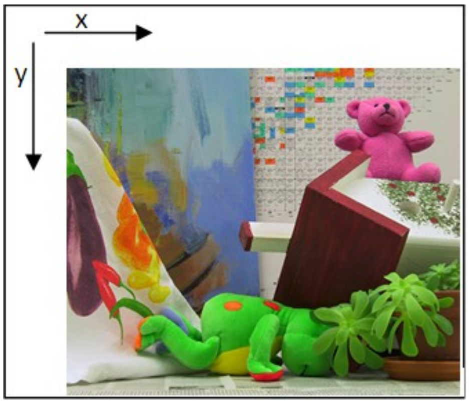


Fig. 5.1. “Teddy” image [1] and coordinate axis.

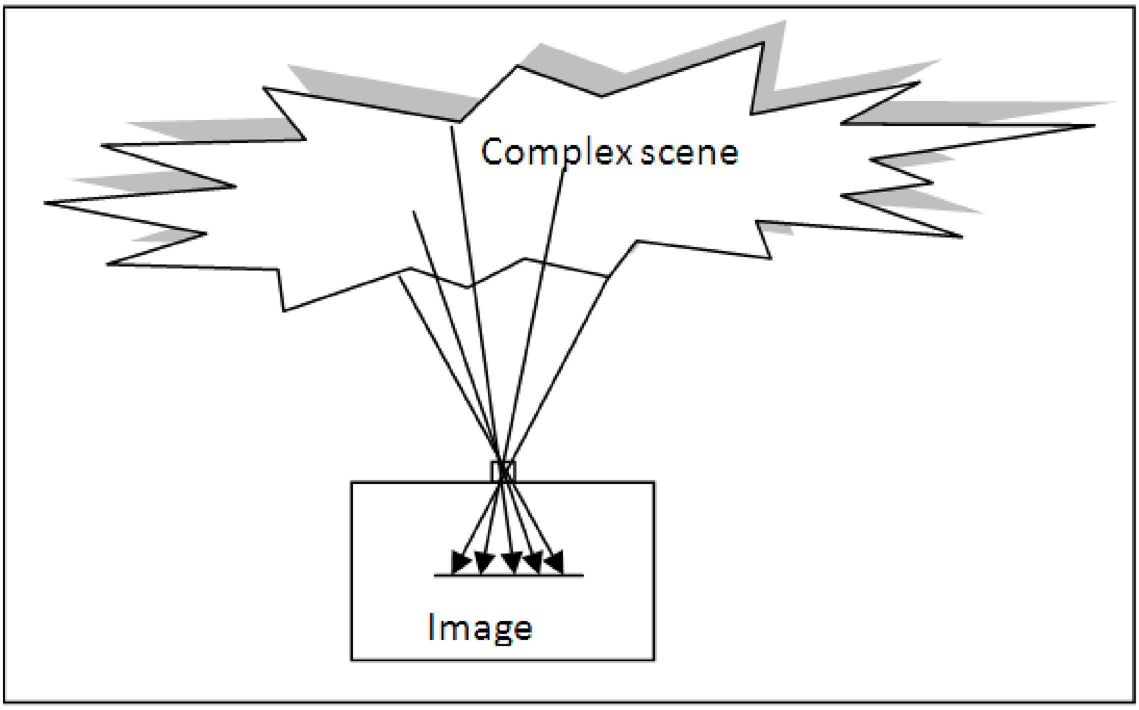


Fig. 5.2. Image as a real scene projection. Top view.

An image is a projection of the scene over a perpendicular plane.

Since in each pixel only one point of the real scene is projected, the depth information is mathematically erased during the projection process into the image plane.

The 3D vision processes have as goal the reconstruction of this lost information, and, thus, the distances from each projected point to the image plane. An example of such reconstruction is shown in figure 5.3.

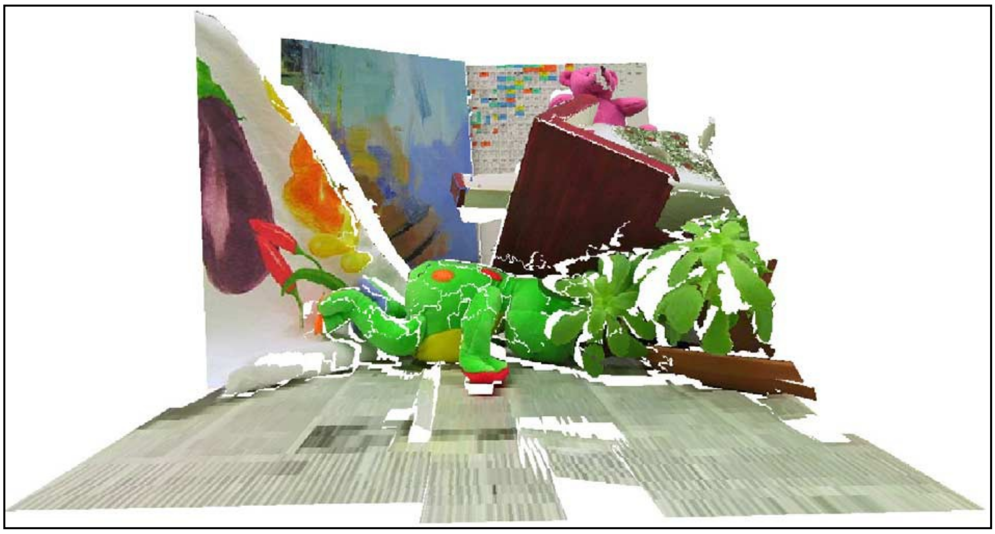


Fig. 5.3. Scene depth reconstruction of image from figure 5.1. [2].

The reconstruction, also called depth map estimation, has to face some fundamental problems:

- On one hand, some extra information has to be obtained, for absolute depth estimation. Topics will be discussed in section 5.3.2.
- On the other hand, there are, virtually, infinite points in the scene that are not projected and, then, must be, in some cases, interpolated. This is the case of occluded points, shown in figure 5.4.

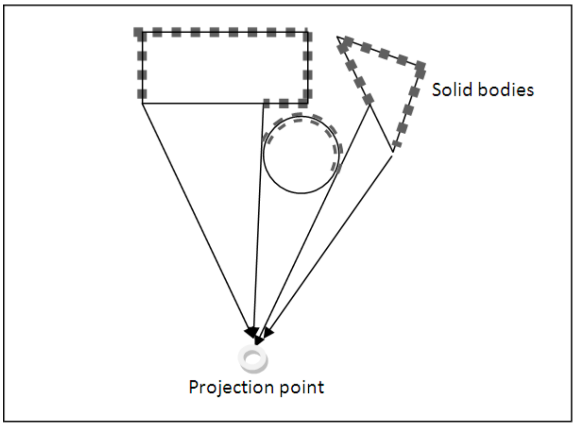
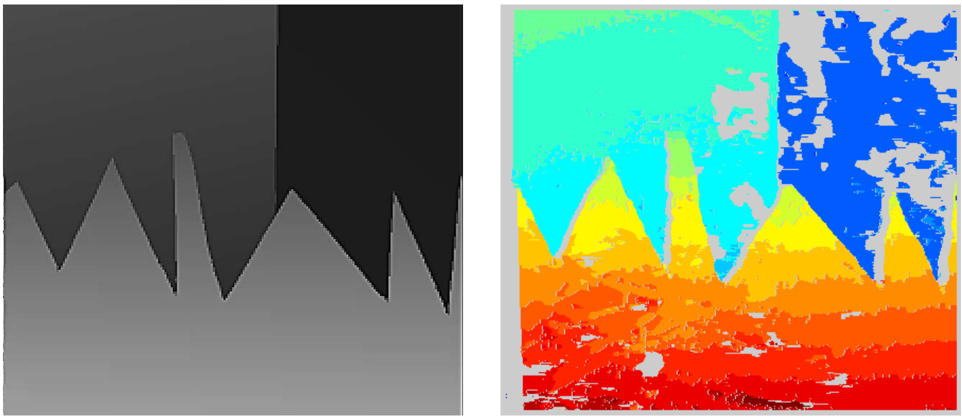


Fig. 5.4. Occluded points, marked in doted contours.

For comparison proposals, it is interesting to represent the depth of a scene in a 2D representation, thus, in an image. These images are the so called 2.5D images, and represent, by means of a parameter variation, the distance of every pixel to the image plane. We can find in literature two main approaches to these representations, shown in figure 5.5.



(a) **(b)**

Fig. 5.5. (a) Gray scale [1] and (b) color [3] representation of the depth map. The first image is the ground truth depth map, while the second one is an estimation.

In the first image, the gray level represents the inverse of the distance. Thus, more a pixel is bright, closer is the point represented. Vice versa, the darker is a pixel, further is the represented point. In the second representation, red-black colors represent closer points, and blue-dark colors the further points. Let's notice that the first representation is the most common one, and that which will be used in this work.

5.2. Human Visual System

It is important to understand how the human visual perception works, regarding the depth perception and image comprehension. The basic structure of human visual system is shown in figure 5.6.

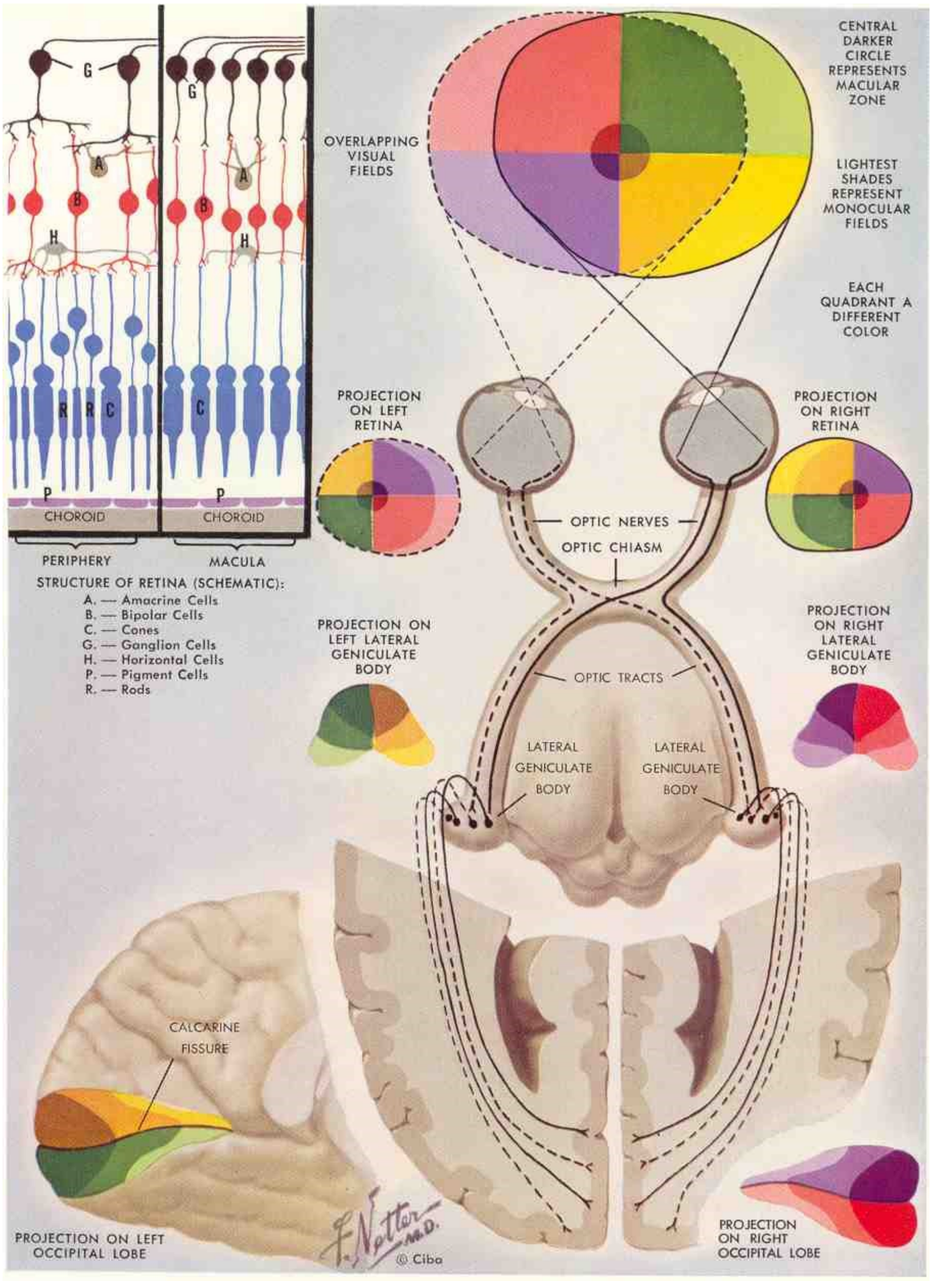

G
A
B
H
R
R
C
C
P
CHOROID
CHOROID
PERIPHERY
MACULA
STRUCTURE OF RETINA (SCHEMATIC):
A. — Amacrine Cells
B. — Bipolar Cells
C. — Cones
G. — Ganglion Cells
H. — Horizontal Cells
P. — Pigment Cells
R. — Rods
OVERLAPPING VISUAL FIELDS
CENTRAL DARKER CIRCLE REPRESENTS MACULAR ZONE
LIGHTEST SHADES REPRESENT MONOCULAR FIELDS
EACH QUADRANT A DIFFERENT COLOR
PROJECTION ON LEFT RETINA
PROJECTION ON RIGHT RETINA
OPTIC NERVES
OPTIC CHIASM
PROJECTION ON LEFT LATERAL GENICULATE BODY
PROJECTION ON RIGHT LATERAL GENICULATE BODY
OPTIC TRACTS
LATERAL GENICULATE BODY
LATERAL GENICULATE BODY
CALCARINE FISSURE
PROJECTION ON LEFT OCCIPITAL LOBE
PROJECTION ON RIGHT OCCIPITAL LOBE
F. Netter M.D.
© Ciba

Fig. 5.6. Schematic human visual system, retrieved from [4].

Depth perception is mainly processed by means of the so-called stereo vision. Stereo vision is the process of retrieving the depth information comparing two images of the same scene taken from slightly different positions.

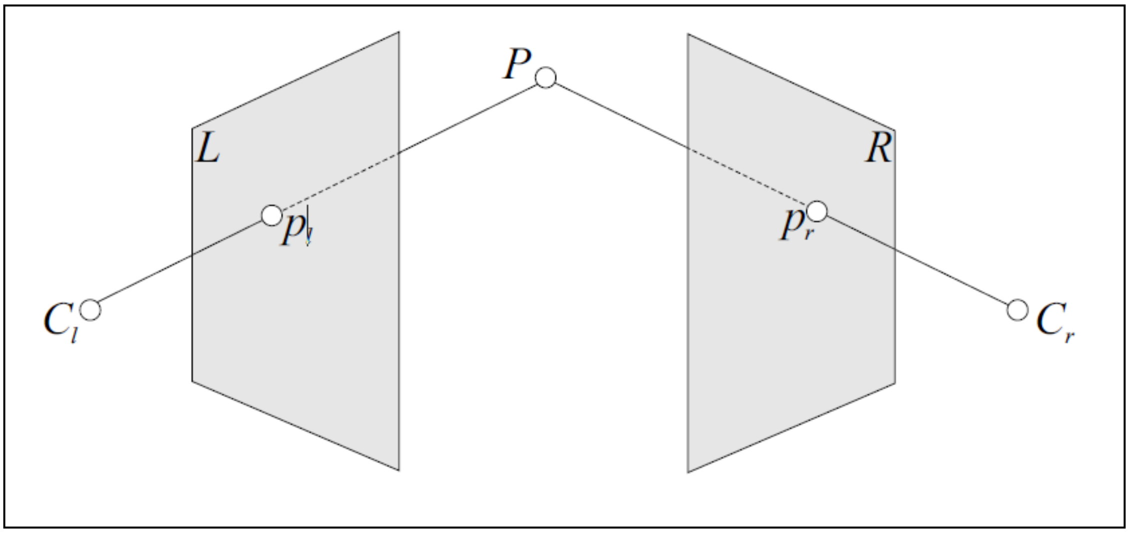
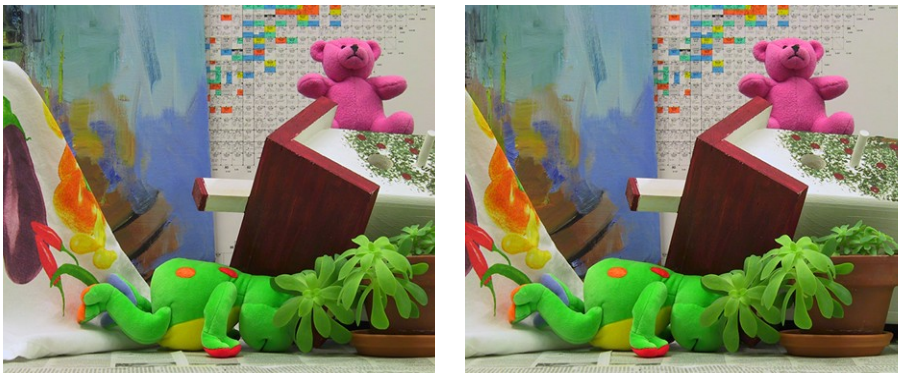


Fig. 5.7. Epipolar geometry of a stereo vision system [5].

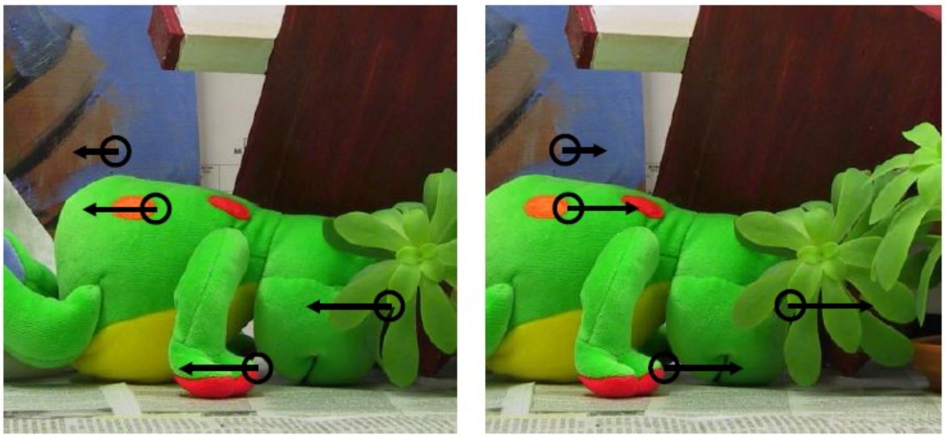
In this figure, C_l and C_r are the focal centers of each sensor, and L and R, the image plane. Both of them are capturing a common point P from a scene, producing p_l and p_r as projections of P over each plane.

We call the epipolar line in the left image plane the set of points in such plane that produces only one projection point in the right image plane, and vice versa.

When some geometrical assumptions are taken into account (which will be discussed in section 5.3.2.2.1), the result of this geometry is a pair of images horizontally decaled.



(a) (b)



(c) (d)

Fig. 5.8. (a) and (b): Stereo pair of images [1], and (c) and (d) corresponding points, decaled regarding the opposite image [5].

In the human visual system, this gap is avoided by moving the eyes to converge to the focusing plane. This gap is, then, null in this plane. This is the situation of figure 5.7.

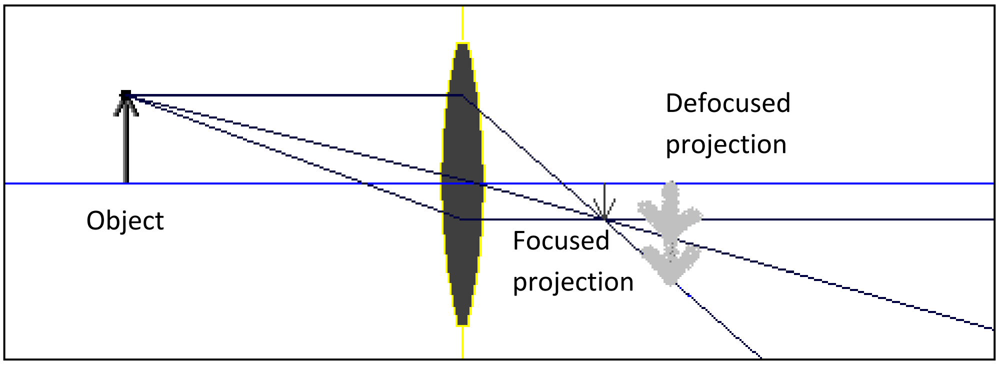
An important psychological task to understand what we see is the segmentation of the image. This procedure splits an image in different regions, without overlap among them, allowing understanding and separating figure from background. In mathematical terms, let Ω be the image domain, thus the segmented regions may be expressed as:

$$\Omega = \sum_{k=1}^{K} S_k \tag{5.1}$$

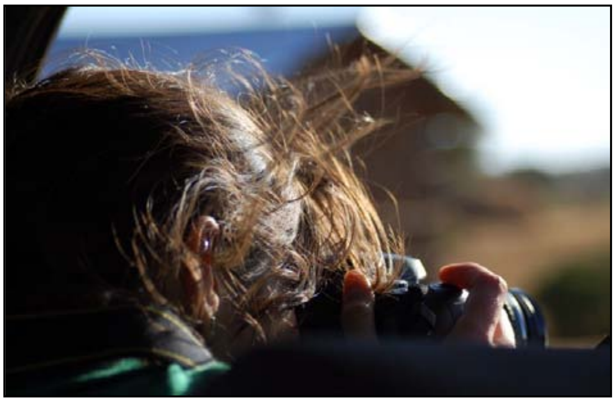
where S_k refers to the k_{th} region and $S_k \cap S_j = \emptyset$ for $k \neq j$ [6].

The human brain applies this technique in a straightforward way. Regarding this point, image segmentation (that is, separate the pixels which belong to an object and those which belong to the background) is a complicated process, and for the human vision, this process has not been well established. Several studies on this issue have been developed [7-9], where the human vision process has been proposed as a multifunctional system where shapes [10], areas [11], colors [12], movements [13] and other visual or psychological characteristics [14] or a mixture of them [15] are involved in the final cognitive separation of the object from the background. This process is helped by means of movement [16] or pattern [17].

Another source of this information for human visual system is the focus. Since the aperture of a sensor is finite and not null, not every point in the projection is focused. This effect, applicable to both human and synthetic visual systems, produces a Gaussian blur on the projected image, proportional to the distance of that point to the focused plane:



(a)

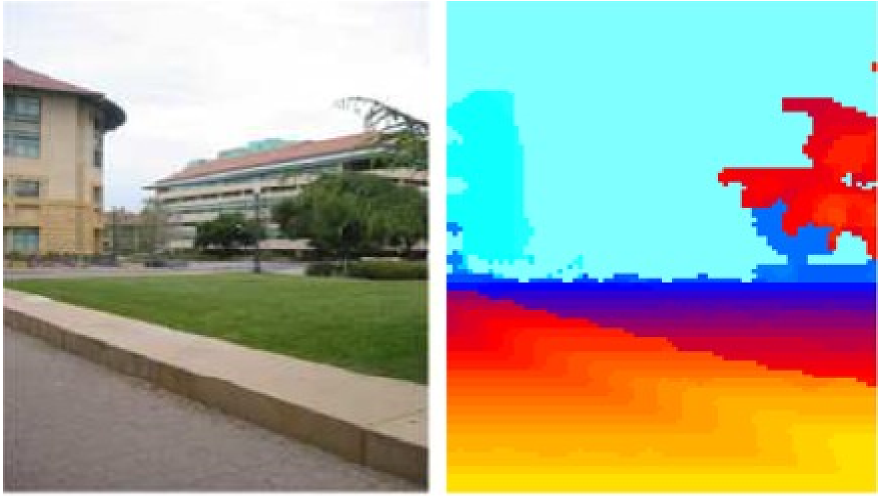


(b)

Fig. 5.9. (a) Focus and defocus scheme and (b) example.

The problem with that approximation is the symmetric effect of defocusing. We cannot know whether an object is closer or farther to the camera from a defocusing measurement.

Finally, another important feature that gives indirect depth information is the structure of the objects appearing in the image, as we can see in figure 5.10.



(a) (b)

Fig. 5.10. (a) A “perspective” image and (b) its ground truth [18]. Notice that nomenclature of colors regarding the measurement of distance changes from image 5.5.

We can appreciate in this example how our brain can reconstruct a relative measurement of the distance by means of the structure of the objects, with some cognitive information about them.

Merging all these features, human visual system produces a very accurate depth measures in short distances (error grows exponentially with the distance).

5.3. State of the Art in 3D vision

Several researches have been published in papers around 3D computation since several years ago, producing a huge amount of proposals and algorithms dealing with different aspects of the 3D vision.

For a general classification, we can divide all proposals as active and passive approximation to the depth map estimation problem.

5.3.1. Active Methods

Active methods have an emitter of some kind of waves, which reflect in the object, allowing computing distances in terms of phase, delay, intensity or other wave parameters. This approximation uses to obtain very accurate depth maps (sometimes it is used to obtain the so called depth or ground truth) as done in [18] and shown in figure 5.10.

Within the field of active methods, several systems are shown which make use of technologies based on laser [18], ultrasound [19], pattern projection [20] or X rays [21].

The main disadvantage of such proposal is the amount of energy needed to measure the distance and the weight of the system. For example, in the case of [18], the laser used works at 20W and weights 4.5Kg.

In some cases, the complexity of image sensors (in ultrasound systems, for example) is also a constraint to build a portable and low-cost device.

This approximation to the 3D vision problem is incompatible with our objectives and, hence, discarded.

5.3.2. Passive Methods

Passive methods allow computing the depth map from a measurement of the natural scene, without any kind of artificial illumination. Generally, they can be implemented with standard and commercial hardware, reducing dramatically the cost of the device. Another advantage is the small amount of energy, regarding the active methods, needed to compute the depth map.

We will divide the approaches in terms of the number of cameras (i.e. images) needed to extract the depth map, as monocular, stereo and multiview vision systems.

5.3.2.1. Monocular Systems

There are several approaches which are based on monocular vision, for example, using structures within the image [22]. In this approximation, some basic structures are assumed, producing a relative volume computation of objects represented in an image. Two examples of such family of algorithms are shown in figure 5.11.

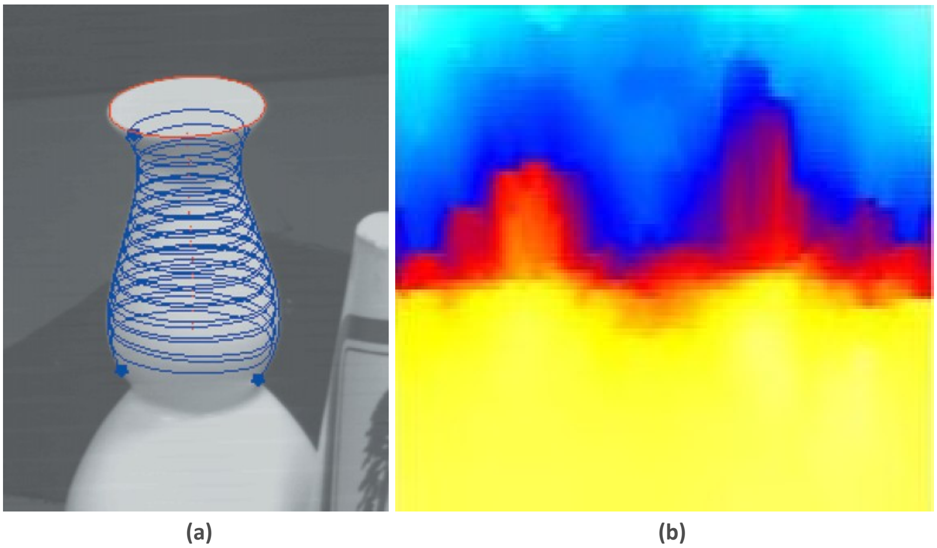


Fig. 5.11. (a) Structure based volume estimation in [22] and (b) a monocular depth estimation from [18].

As explained in the referenced works, the measurement of distances in this proposal is relative. We cannot know the exact distance to each point of the image but just the relative distance among them. Moreover, some other disadvantages of these algorithms arise from the intrinsic limitation in terms of expected forms and geometries of figures appearing in the image. Perspective can trick this kind of algorithms producing uncontrolled results. Finally, they are very dependent on image noise, which can alter the initial segmentation of objects.

More accurate and flexible approximations are based on the relative movement of independent points located in the image:

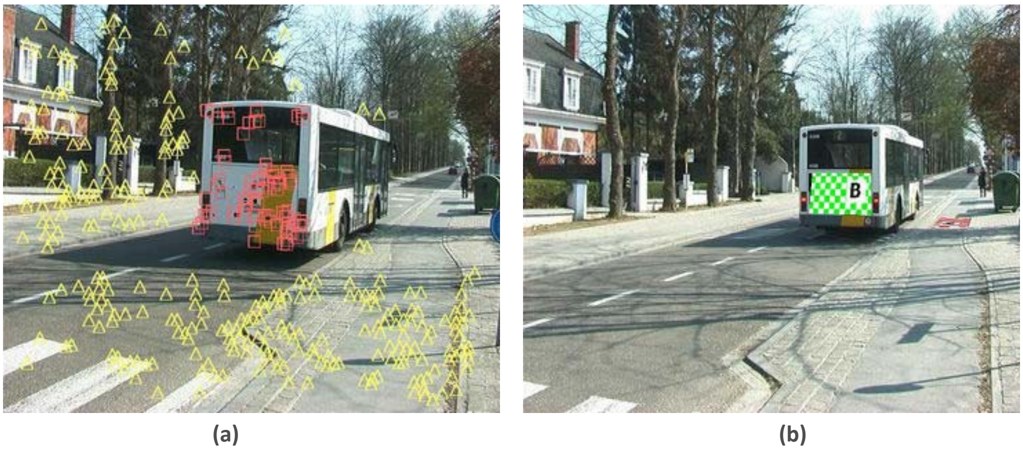
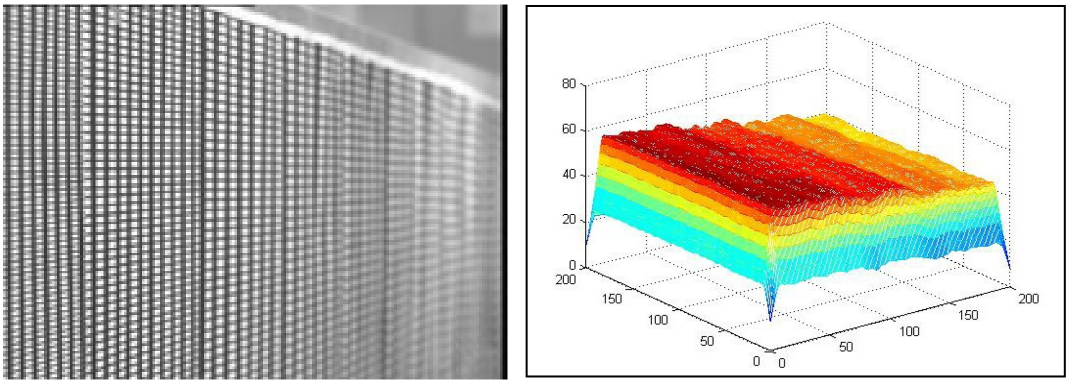


Fig. 5.12. Augmented reality and 3D estimation through points relative movements in [23].

The main features of these algorithms are the dependency of a video sequence (to compute the optical flow of descriptor points) and the relative measurement of the distance.

The only approach that provides an absolute measurement of distance with monocular information is based in the focus properties of the image. This approach estimates the distance of every point in the image by computing the defocusing level of such points, following the human visual focusing system. This defocusing measurement is mainly done with Laplacian operators, which computes the second spatial derivative for every point in a neighborhood of N pixels in each direction. Many other operators have been proposed, and a review of them can be found in [24].

Focused pixels provide an exact measurement of the distance, if the camera optical properties are known.



(a) (b)

Fig. 5.13. Planar object distance estimation by focus [25].

This approximation has important errors when defocusing is high, and is very sensitive to texture features of the image and other noise distortions.

5.3.2.2. Stereo and Multiview Systems

Monocular systems present important constraints and limitations to depth map estimations, regarding the objectives of this thesis work. The only absolute measurement is achieved sacrificing accuracy in the closer and further objects, and in real world images errors due to textured objects are very high.

That is the main reason why some other proposals have been presented for 3D vision. These approaches are based on multiview 3D systems. The main idea of this proposal is the use of two or more images of the scene taken from separated positions. Following the human visual system, the stereo vision approach uses two cameras ([26] for example). Multivision systems attempt to compute the depth map by recomposing the scenario from different points of view [27, 28], or obtaining images from a broad range of angles [29]. Processing two or more images produce new effects that must be taken into account.

5.3.2.2.1. Epipolar Geometry and Image Rectification

The structure shown in figure 5.14 represents the generalized epipolar geometry problem. Two cameras pointing in an arbitrary angle to a scene where there are some common points projected in each image planes.

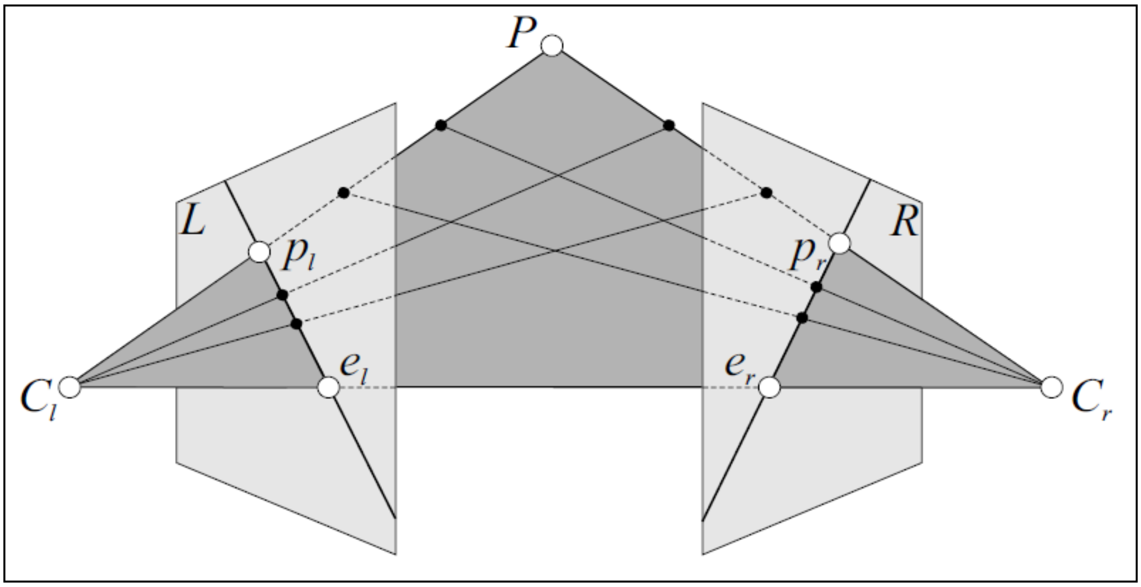


Fig. 5.14. Generalized epipolar geometry problem [5].

Data collected from these systems present a high amount of geometric distortions that must be corrected [30].

This problem can be simplified by a physical/mathematical rotation of the cameras/images to perceive the scene from the same tilt angle and pointing to infinite, as shown in figure 5.15, or by means of geometrical processing of the image.

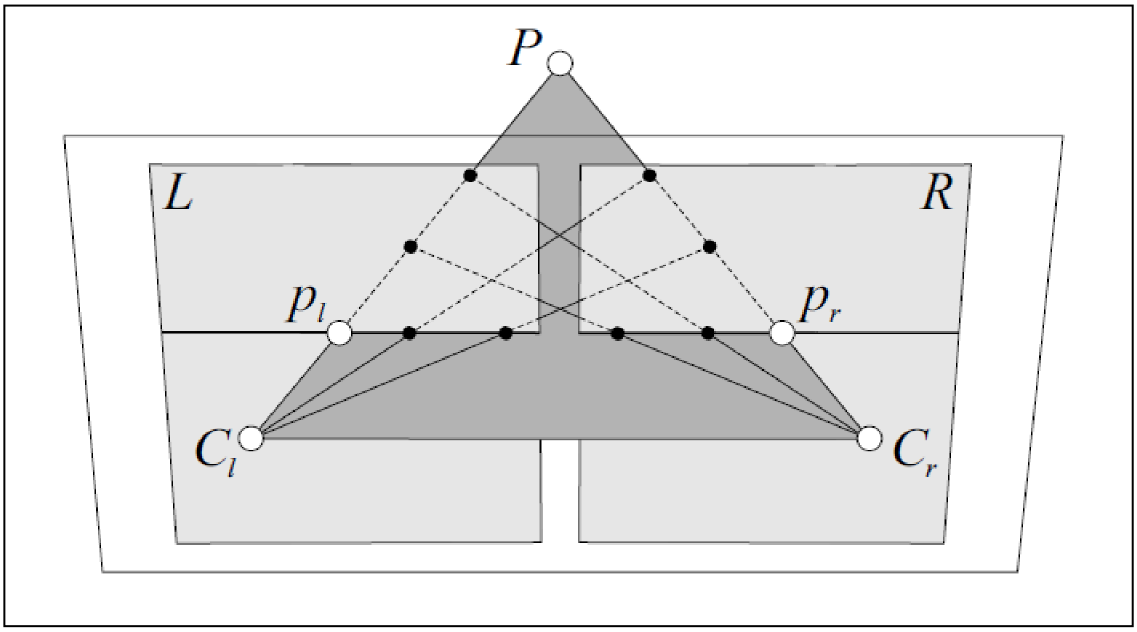


Fig. 5.15. Fronto-parallel simplification of the generalized epipolar problem [5].

These assumptions are described in detail in [27, 31] and are referred to as fronto-parallel hypothesis. The result is the so called "rectified images".

In this approximation, there is only one epipolar line in both images, so we can assume that:

- y_r=y_l
- Distance of a point = $K/(x_l\text{-}x_r)$

where $y_{l,r}$ and $x_{l,r}$ are the ordinate and abscissa coordinates of the left and right points.

The height of every common point is the same in both images. Moreover, the distance to the sensor will be inversely proportional to the difference of the abscissa values of the projected points, being the left projection abscissa higher than the right one (the zero is referred to the left side of the image). A point in the infinite will have equal abscissa coordinate in both projections. This assumption is generally used [32].

Another aspect to be considered is the fact that the left part of the left image is not captured in the right image, and vice versa, as shown in figure 5.16.

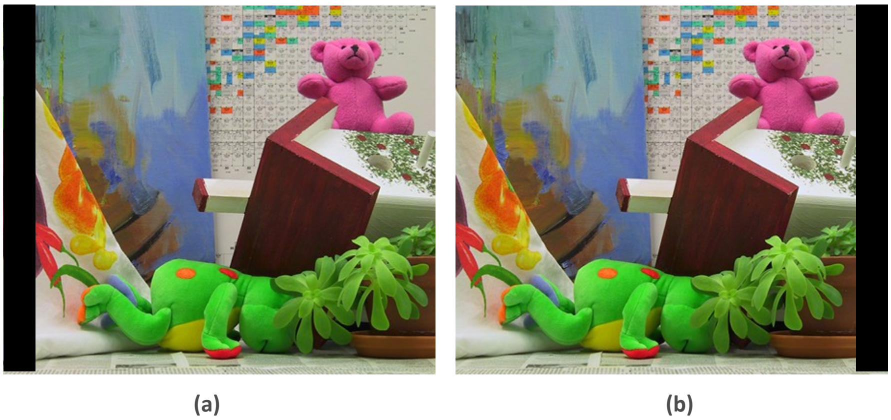


Fig. 5.16. Teddy image pair, with shadowed areas that are not shared.

As a final consequence of this hypothesis, let's remark that there is a minimum distance measurement possibility.

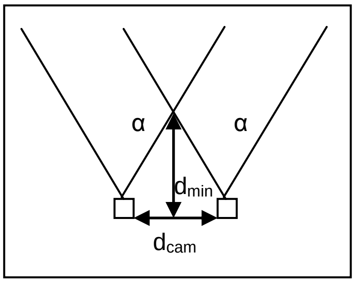


Fig. 5.17. Minimum distance measurement constraint.

In this graph, d_{min} is the minimum distance an object can be from the surface of the camera dock to allow the depth map reconstruction, in terms of the optical aperture of the cameras α and the distance between cameras d_{cam}.

5.3.2.2.2. Matching

When different viewpoints from the same scene are compared, a further problem arises that is associated with the mutual identification of images. The solution to this problem is commonly referred to as matching. The matching process consists of identifying each physical point within different images [27]. However, matching techniques are not only used in stereo or multivision procedures but also widely used for image retrieval [33] or fingerprint identification [34] where it is important to allow rotational and scalar distortions [35].

In stereo vision, the difference observed between images is referred to as disparity and allows retrieving depth information from either a sequence of images or from a group of static images from different viewpoints.

There are also various constraints that are generally satisfied by true matches thus simplifying the depth estimation algorithm, these are: similarity, smoothness, ordering and uniqueness [2].

This approach solves the problem with four main strategies: local, cooperative, dynamic programming and global approximations.

The first option takes into account only disparities within a finite window or neighborhood which presents similar intensities in both images [36, 37]. The value of a matching criterion (sum-of-absolute-differences (SAD), sum-of-squared-differences (SSD) or any other characterization of the neighborhood of a pixel) in one image is compared with the value computed in the other image for a running window. These windows are k×k pixel size. Then, this sum is optimized and the best match pixel is found. Finally, the disparity is computed from the abscissa difference of matched windows:

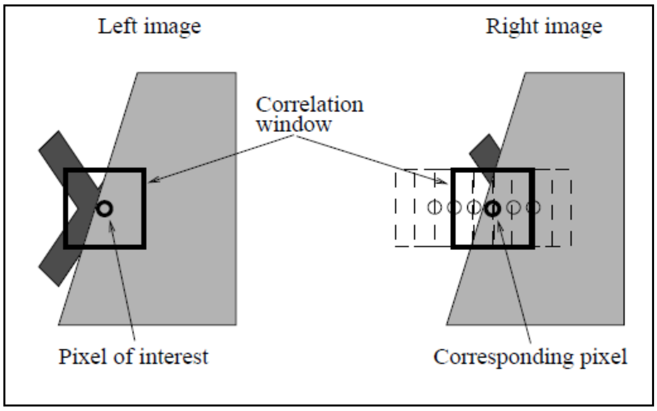


Fig. 5.18. Moving window finding an edge. Graph taken from [38].

Results of this kind of algorithms are not so accurate. In figure 5.18 we show a typical result of these algorithms.

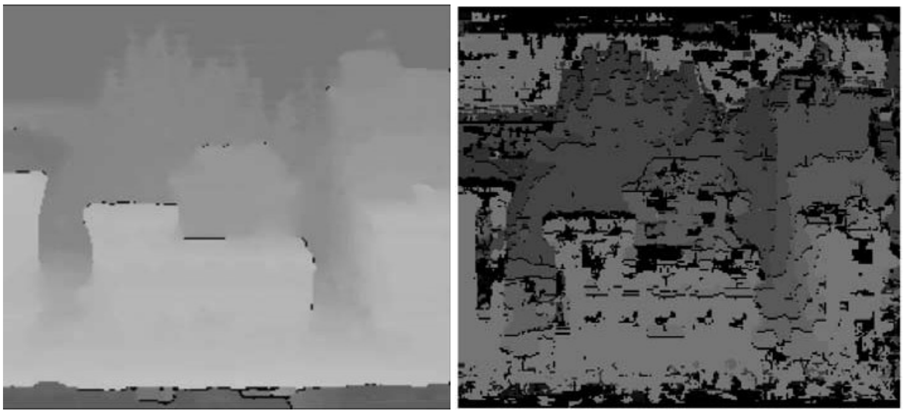

(a) (b)

Fig. 5.18. (a) Very accurate depth map, compared to (b) window-based disparity estimation [39] (right).

This result is obtained with a 7×7 matching window. The main disadvantage can be clearly seen: the number of operations needed gives a global order of the algorithm of $o(n)=N^3 \cdot k^4$ for a N×N image with windows of k×k pixels. This order is very high and these algorithms not so fast, around 1 and 5fps [38] the fastest one.

Another possibility for local matching is implemented by means of point matching. The basic idea consists on identifying important points (information relevant) in both images and matching the features between them:

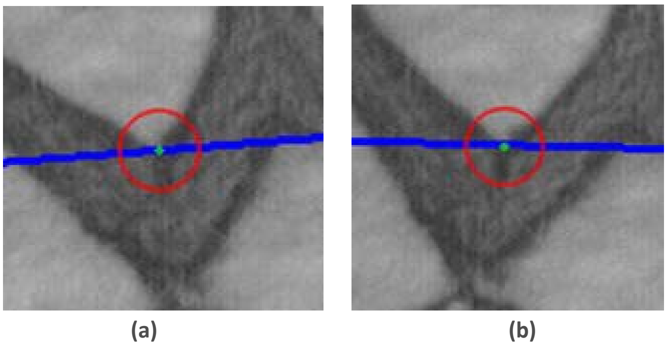


Fig. 5.19. Relevant points retrieval. In blue, epipolar lines in each image plane. Taken from [40].

After this process, all relevant points are identified, as shown in figure 5.20.

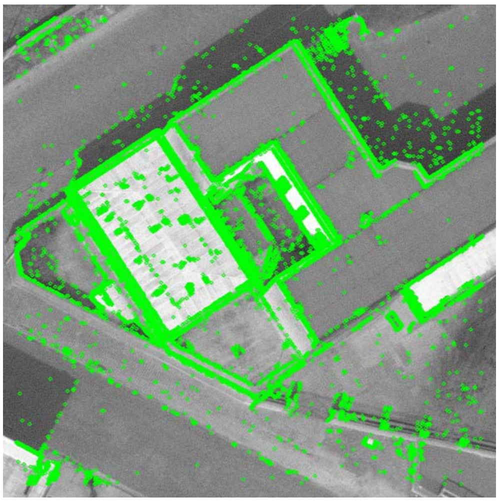
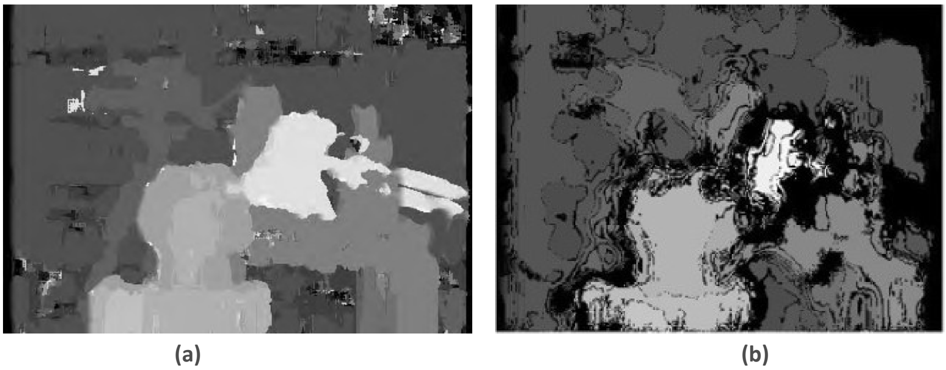
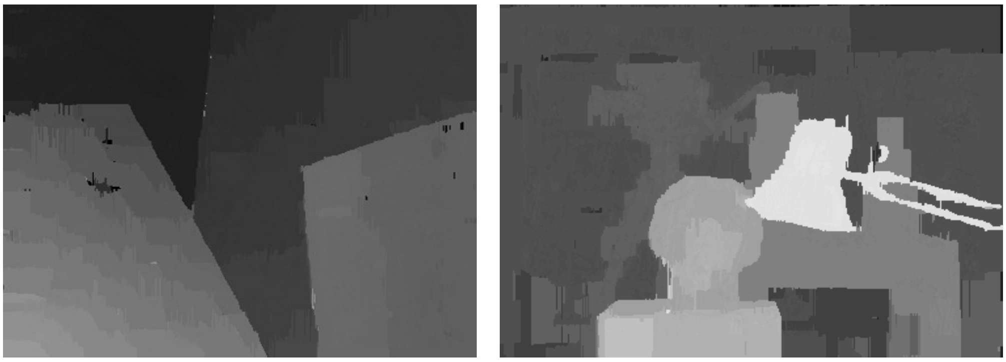


Fig. 5.20. Relevant points extracted [40].

Results of these algorithms are more accurate than those presented previously as shown in Fig. 5.21.





(c) (d)

Fig. 5.21. Two results with the Tsukuba pair: (a) from [41] and (b) from [42]. (c) and (d) results for Venus and Tsukuba pair of images from [43].

These algorithms are neither too fast, achieving processing times of few seconds [44]. In the case of Liu, he gives time measures to obtain these results with a Pentium IV (@2.4GHz): 11.1 seconds and 4.4 seconds for the Venus and the Tsukuba pairs respectively.

The main drawback is the necessity of interpolation. Only matched points are measured. After that, an interpolation of the non-identified points is mandatory, increasing slightly the processing time. Another important disadvantage is the disparity computation on untextured surfaces.

Cooperative algorithms were firstly proposed by Marr & Poggio [45] and it was implemented trying to simulate how the human brain works. A two dimensional neural network iterates with inhibitory and excitatory connections until a stable state is reached. Later, some other proposals in this group have been proposed [46, 47].

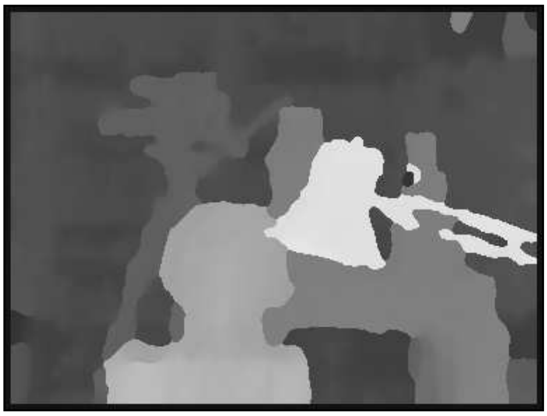


Fig. 5.22. Tsukuba depth map computed with a cooperative algorithm [47].

Global algorithms make explicit smoothness assumptions converting the problem in an optimization one. They seek a disparity assignment that minimizes a global cost or energy function that combines data and weighted (with λ) smoothness terms [32, 48] in terms of the disparity function *d*:

$$E(d)=E_{data}(d)+ \lambda \cdot E_{smooth}(d) \quad (5.2)$$

Dynamic programming strategy consists on assuming the ordering constraint as always true [48]. The scanline is assumed, then, to be horizontal and unidimensional. The independent match of horizontal lines produces horizontal "streaks". The problem with the noise sensitivity of this proposal is smoothed with vertical edges [49] or ground control points [50].

Some of the best results with global strategies have been achieved with the so-called graph cuts matching. Graph cuts extend the 1D formulation of dynamic programming approach to 2D, assuming a *local coherence constraint*, i.e. for each pixel, neighborhoods have similar disparity. Each match is taken as a node and forced to fit in a disparity plane, connected to their neighbors by *disparity edges* and *occlusion edges*, adding a source node (with lower disparity) and a sink node (highest disparity) connected to all nodes. Costs are assigned to matches, and mean values of such costs to edges. Finally, we compute a minimum cut on the graph, separating nodes in two groups and the largest disparity that connect a node to the source is assigned to each pixel [48]. Figure 5.23 presents an example of these algorithms.

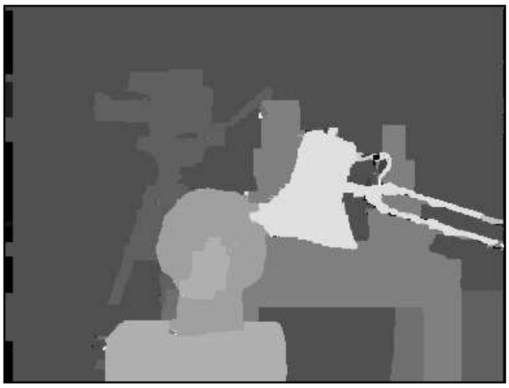


Fig. 5.23. Graph cuts depth estimation [51].

We can find also a group using some specific features of the image, like edges, shapes and curves [33, 52, 53]. In this family, a differential operator must be used (typically Laplacian or Laplacian of Gaussian, as in [26, 54]).

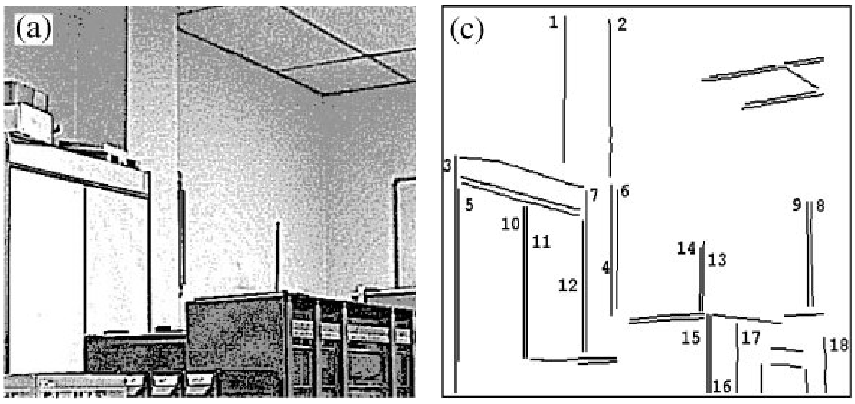


Fig. 5.24. Laplacian filtering to extract edges. Images taken from [54].

This task requires a convolution of 3×3, 5×5 or even bigger windows; as a result, the computing load increases with the size of the operator (for separable implementations). However, these algorithms allow real-time implementations:

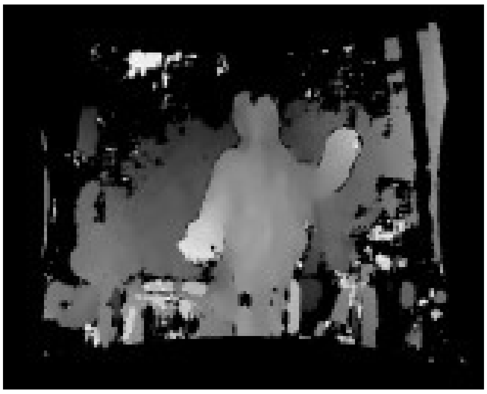


Fig. 5.24. Depth map estimation through differential operator, from [26].

The speed of the algorithm presented in [26] needs around 30ms over FPGA hardware.

Another possibility of global algorithms are those of Belief propagation [55], modeling smoothness, discontinuities and occlusions with three Markov Random Fields and itinerates finding the best solution of a "Maximum A Posteriori" (MAP). An example of this algorithm can be found in figure 5.25.

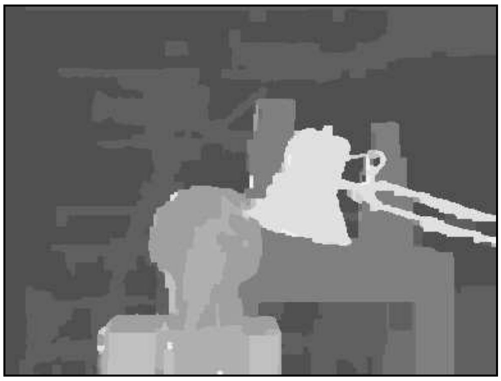


Fig. 5.25. Belief propagation depth map estimation [55].

A final family of global algorithms to be studied in this state of the art is the segment-based algorithms. This group of algorithms chops the image as explained in equation 5.1 to match regions. An initial pair of images is smoothed and segmented in regions:

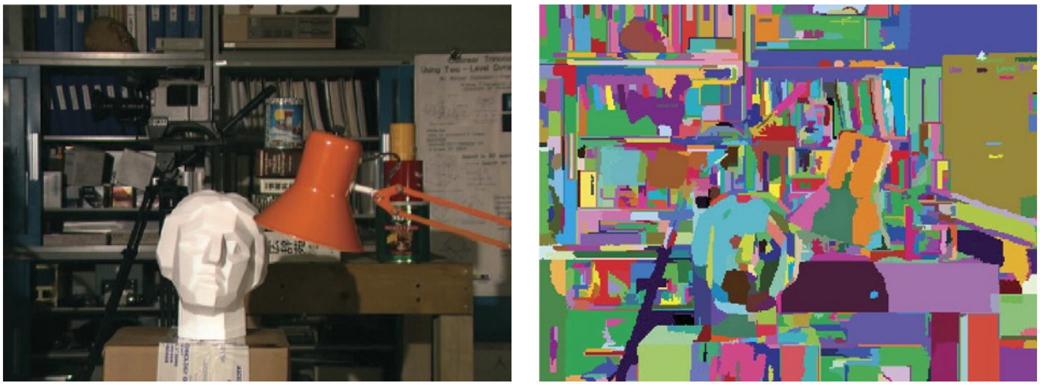


Fig. 5.26. Image segmentation of the Tsukuba pair of images. Left, original image. Right, segmented version. Taken from [2].

The aim of this family of algorithms focuses the problem of untextured regions.

After forcing pixels to fit in a disparity plane, the depth map estimation results in the following one:

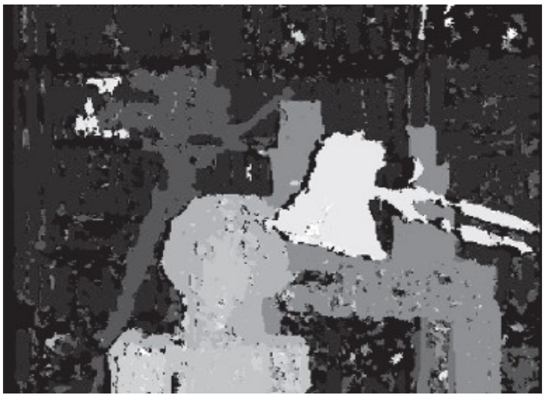


Fig. 5.27. Initial disparity map with segment regions and plane fit forcing. Image extracted from [2].

These algorithms have the advantage of producing a *dense depth map*, disparity estimate at each pixel [32], thus, avoiding interpolation. Presented algorithm also perform a k×k window pre-match, and a plane fitting, producing a high computational load (and computation time of tens of seconds), and avoiding its use in real-time applications [2].

Combinations of segment-based and graph cuts algorithms have also been implemented [56].

A further group of algorithms are based on wavelets, as described in [53]. These algorithms present similar problems in terms of time performance.

To sum up, each of the previously described approaches to matching problem presents several computing problems. In the case of edges, curves and shapes, differential operators increase the order linearly with the size (for separable implementations). This problem gets worse when using area-based matching algorithms, following the computational load an exponential law. The use of a window to analyze and compare different regions is seen to perform satisfactorily [2] however this technique requires many computational resources. Even most of segment-based matching algorithms perform a N×N local windowing matching as a step of the final depth map computation [32, 56]. It is important to notice that this step is not dimensional separable. Most of these algorithms, however, obtain very accurate results, with the counterpart of interpolating optimized planes that forces to solve linear systems [56, 57]. The calculations required for depth mapping of images is very high. It has been studied in detail, and a complete review of algorithms performing this task by means of stereovision can be found at [32].

5.4. Contributions

As explained before, our aim is a low-cost algorithm, implementable in cheap, light and standard hardware and with very low power consumption. With those constraints, we can sacrifice accuracy in order to obtain real-time approximations.

This work proposes 4 options to overcome some of the found problems in systems reported in the literature, regarding, as a priority, the computation speed over standard PC in C code. The implementation was made using the OpenCV open computer vision library, from Intel.

5.4.1. Segment-based Matching using Characteristic Vectors

With this algorithm, I propose a novel matching algorithm based on characteristic vectors for gray scale images. The vectors are extracted with a region growing algorithm.

5.4.1.1. Characteristic Vector Extraction Algorithm

In order to do so, a loop is called for each region, which starts at the seed pixel. The algorithm verifies that the intensity of the pixels adjacent to the position of the reference pixel intensity is within a pre-defined interval of the reference. If this is the case, the same algorithm is launched on each pixel that complies with the previous condition. Another approach is based on edge information contained within the image [6]. Up until now, this has been considered as the principle definition for any region growing algorithm.

The main problem associated with these algorithms is the requirement for human interaction to correctly place the seeds. Several semiautomatic region growing algorithms have been proposed, where some of these use random seeding [58], manual [59] or semiautomatic seeding [60] and others exist which are based on a completely automatic seeding process [61].

To solve the seed problem, the proposal described in [61] has been followed: whenever a new pixel that does not correspond to the previous region is found (i.e. does not match the inclusion condition) it is marked as a seed. The matching condition is implemented over a dynamic threshold, explained in B. Seeds will grow when the previous region has been fully segmented. A global constraint has been implemented which limits only one growing seed per region. The following pseudo code explains how the algorithm operates:

```
Set up-left pixel as seed
While remains regions (RR) to be analyzed do
  While remains pixels (RP) in the current region do
    Calculate new threshold
      If current pixel is not processed Then
        Compute characteristics (see table I)
          If some of up/down/left/right pixels are not processed Then
            If they are compliant with the threshold Then
              Store their coordinates to be processed in RP
            Else
              Convert it as seed
  If region_Area < minimum_Area Then
    Reset characteristics and index
  Update region characteristics and global variables for new region segmentation.
```

A further issue regarding this type of algorithm must also be taken into consideration. When a synthetic and simple image is analyzed in regions where there are homogenous and heterogeneous gray levels, no problems regarding boundary detection arise. However, in real images problems occur. Regions in gray-scale images are not determined canonically. Thus, a threshold and a reference, which determine whether a pixel belongs or not to the actual region, must be defined.

In this proposal, the threshold is static, but the reference value is dynamically computed in each iteration (if current pixel complains with the inclusion condition), taking into account previous pixels already processed:

$$R_i = \frac{p_i + H \cdot R_{i-1}}{H+1} \tag{5.3}$$

where *H* is a constant weighting the previous reference *R_{i-1}* and *p_i* is the current pixel value.

For every iteration, the inclusion condition is finally computed as follows:

$$\left|R_{i-1} - p_i\right| \leq Th \tag{3.4}$$

where *Th* is the static threshold set at the startup.

This task affects to the quality of the segmentation: techniques regarding the implementation of an objective function to evaluate the quality of the segmentation process have been discussed by other authors (see revision [62]). In an unpublished work, this research propose the use of the weighting factor H as 10 to obtain the reference and a static threshold of 12 computed using the Moran's I objective function, described in [63], applied to a Montecarlo simulation and optimization process. It must be pointed out here that this particular aspect has not been considered as crucial to attain the principle goal of the work presented in this section. The quality of the segmentation process is not the point in this proposal and, thus, will not be taken into account.

The characteristics of an object or region which has been separated from the rest of the image are useful for the post-processing of independent parts of an image, i.e. identification, modification, labeling, compression etc. These characteristics must be efficiently measured and optimized since the region characteristics are computationally costly, such as border identification, spatial median filter, etc. These useful but computational costly characteristics are, therefore, one of the main drawbacks for achieving a real-time system.

In this section, we propose a set of characteristics extracted on-the-flight while the region growing algorithm is operating. The basic idea here consists of extracting the characteristics, which define the object where independent pixel operations instead of regional, are performed. In this situation each pixel is only analyzed once, which is the minimum number of operation that must be performed for an image analysis. As a result of this limitation, all of the characteristics may not be obtained. In table 5.1, the characteristics extracted from this algorithm are shown as well as how it operates on the i-th step.

Static Characteristics	Formula in the i-th step [condition of application]
Intensity	$value(pixel_i)$
Frontier length (if applicable to i-th pixel)	$F_i = F_{i-1} + 1$
N-M Moment	$M_i^{nm} = M_{i-1}^{nm} + x^n y^m$

Table 5.1. On-the-flight object characterization.

As is shown in this table, the intensity of the pixel with tolerance ε has been included in the basic characteristics, as a sine qua non condition for correct pixel identification and segmentation. Another basic feature that permits a well defined characteristic vector is the frontier length.

The N-M moment is a spatial descriptor of the region, and considers all the pixels to which it is applied, where all of these have the same value. With this mathematical tool, several characteristics of interest may be obtained [64] as follows:

Area:

$$A = M_{00} \tag{5.5}$$

Centroid:

$$x_c = \frac{M_{10}}{M_{00}}, y_c = \frac{M_{01}}{M_{00}} \tag{5.6}$$

Let $a = \frac{M_{20}}{M_{00}} - x_c^{\,2}, b = \frac{M_{02}}{M_{00}} - y_c^{\,2}$ and

$$c = 2\left(\frac{M_{11}}{M_{00}} - x_c y_c\right),$$

Then, the orientation is:

$$\theta = \frac{\arctan\left(\frac{c}{a-b}\right)}{2} \tag{5.7}$$

This last characteristic is an important guideline regarding the human vision systems ability to recognize objects [65].

Final spatial descriptors are computed as shown bellow [64]:

Length of the region:

$$l = \sqrt{\frac{(a+b) + \sqrt{c^2 + (a-b)^2}}{2}} \tag{5.8}$$

Width of the region:

$$w = \sqrt{\frac{(a+b) - \sqrt{c^2 + (a-b)^2}}{2}} \tag{5.9}$$

These characteristics are the first two eigenvalues of the probability distribution which represent the region.

As may be seen, 6 different moments must be computed for each pixel: M_{00}, M_{10}, M_{01}, M_{20}, M_{02} and M_{11}. After the algorithm has operated on the complete region, all the moments described above are available thus the characteristics pertaining to equations (5.5-5.9) can be obtained.

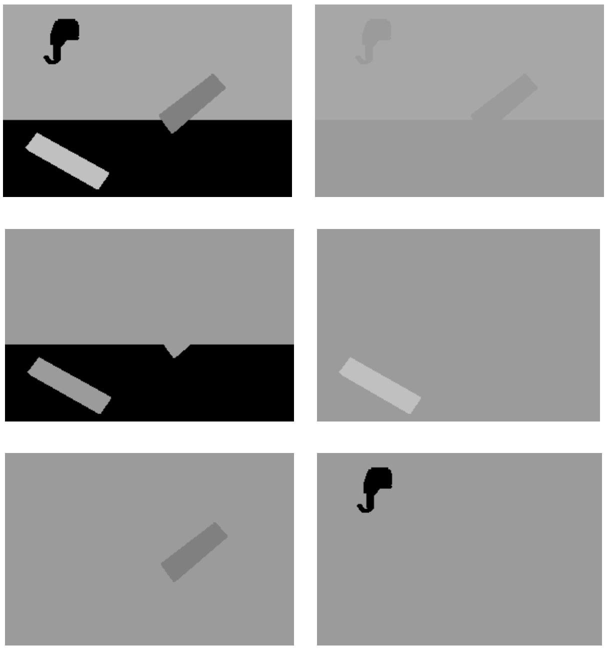


Fig. 5.28. Synthetic test: Original synthetic image (up left) and the extracted regions. The extracted regions are shown with their original gray scale level on 155 gray level background.

	Region (up-right)	Region (middel-left)	Region (middle-right)	Region (bottom-left)	Region (bottom-right)
Area [pixels]	33701	21946	1841	1637	875
xc (of centroid) [pixels]	149.7	156.2	66.6	195.3	
yc (of centroid) [pixels]	58.5	159.6	162.4	102.36	33.4
Angle [º]	1.2	-0.07	-28.9	39.1	
Length	87.7	86.8	25.1	21.1	13.7
Width	34.4	23.6	6.1	6.5	
Frontier [pixels]	662	620	292	262	188

Table 5.2. Characteristics extracted from figure 5.28.

This processing, implemented over OpenCV library and on a 1,6GHz microprocessor lasts 11.8ms, thus, 196ns per pixel.

This algorithm was described and published in the paper [66].

5.4.1.2. Gray Scale Stereo Matching Algorithm

Based on the previous presented algorithm, the next step was the implementation of a stereo vision algorithm via characteristic vectors matching.

Considering the stereo matching approach, there are several geometrical and camera response assumptions that have been made to compare two images that are slightly different. These assumptions have already been discussed in previous sections in this chapter.

Taking this into account, a region matching algorithm is proposed, that reduce the number of operations needed for stereo matching, obtaining at the same time results that are relevant compared to those found in the bibliography.

This algorithm works as follows:

1. Image preprocessing. First of all, a Gaussian low pass filter is explained, to reduce outlier pixels that are not representative. This task is crucial to perform the region growing algorithm.
2. Features extraction by region growing. Secondly, the region growing algorithm is applied and regions descriptors are obtained.
3. Vectors Matching. Once the vectors with the extracted descriptors are created, the matching process over the pair of vectors (one vector for each region, one array of vectors for each image) is implemented.
4. Depth Estimation. The depth estimation is computed from the horizontal distance of the centroid of every matched pair of region descriptors.

The application of region growing algorithms is generally preceded by image processing techniques to adjust the visual data to the algorithm affordable range. The proposed algorithm has been designed to operate on grey scale images. Color images are first converted to 256 gray levels by considering their brightness.

Additionally, a smoothing filter is applied to reduce the influence of the noise on the processing. This is carried out using a 3×3 Gaussian filter.

The scope of this algorithm is restricted to the fast segmentation of different regions, and not coherent image segmentation (the fact that different segmented parts belong or not to the same physical object is not of interest). Over-segmentation is then tolerated. An efficient method to set regions is done by truncating the image (and losing some information). In this implementation, this is carried out by using the 3 most significant bits and, then, the gray scale is reduced to 8 levels. The truncation process has been implemented on-the-flight inside the region growing and characteristics extraction algorithm. This algorithm is, then, simplified to a static threshold implementation, saving some operations and memory resources.

The most relevant characteristics obtained from each region are the area, gray value, centroid, length, width, boundary length and orientation, where these have been based on the most relevant visual cues used in human vision [9].

When the segmentation step has been performed, a set of characteristic vectors describing each region of the image is provided. In addition to these vectors, an image is required that maintains the reference between each pixel and the region identifier to which it belongs. This image is referred to as the “footprint image” and it is composed of one byte per pixel which represents the index of the characteristics region identifier in the vector. This fact limits, in this implementation, the number of segmented regions to 255 (the value '0' is reserved for unlabeled and occluded pixels).

Another advantage of this method over the windows and edges matching algorithm is that the interpolation process required to create the disparity surfaces, as reported in [31], is avoided. This is due to the fact that in our approach virtually all pixels are linked to a region (several pixels remain unlabeled in the segmentation process as they belong to regions with insufficient area to be considered). This procedure is then a dense matching.

Using this novel algorithm, a chain of conditions has been proposed to verify the compliance between regions. With this structure, increased efficiency is achieved as every test is not always performed. The majority of the region characteristics are compared according to expression 3.10:

$$val = \frac{abs\left(Ch^{i}_{left} - Ch^{i}_{right}\right)}{\max\left(Ch^{i}_{left}, Ch^{i}_{right}\right)} \tag{5.10}$$

Where *i* represents the *i-th* characteristic (Ch) of those presented in table 1 of the left or right image.

For this case, the possible range of differences is between [0, 1]. We refer to this particular operation as Relative Difference.

Table 5.3 shows the order of conditions, the compared characteristic and the acceptance threshold. Thresholds and comparison tests will be explained after the table.

Item	Characteristic	Comparison test	Acceptance thresholds
1	Centroid ordinates	Absolute difference	<(Image Height)/4
2	Centroid abscissa	Absolute difference and non-negative	[0, (Image Width)/4]
3	Value	Equality	1
4	Area	Relative difference	<55%
5	Length	Relative difference	<30%
6	Width	Relative difference	<30%
7	Angle	Weighed difference	<65
8	Vote of characteristics	Absolute addition	Total Threshold

Table 5.3. Comparison tests and their corresponding thresholds.

The centroid coordinate matching in tests 1 and 2 is only searched in ¼ of the image in each axis, assuming all the potentially matched objects are located far enough from the cameras and in the same scan-line (¼ image size vertical tolerance). Moreover, the difference of left and right centroid abscissas cannot be negative (which should represent objects far away from the infinite, or a myopic orientation of cameras. Both cases are not taken into account since the specified geometrical assumptions are applied).

The preprocessed images, as said before, have been truncated, so pixel values must have the same values to be included within the same region, as done in the third test.

The angle comparison (item number 7) needs a deeper explanation. Because of the ambiguity of the angle measurement when length and width are similar, a comparative function described by equation 3.11 has been implemented:

$$\Delta\alpha = 100 \cdot abs\left(\frac{2}{\pi} \cdot a\tan\left(\min\left(L_l - W_l, L_r - W_r\right)\right) \cdot \sin\left(\alpha_l - \alpha_r\right)\right) \tag{5.11}$$

In equation 5.11 $L_{l,r}$ and $W_{l,r}$ are the length and the width of the left and right image regions, respectively, and $\alpha_{l,r}$, denoted by the subscript l and r are the angles of the left and right image region, respectively. By using this equation the magnitude of the angle is maximized when there is a large difference between the length and width, and vice versa. This is done because the angle measurement of a compact object (similar length and width) is highly noise sensitive and is not suitable as a representative descriptor for the region.

Finally, if the result from all of the previous tests is positive, all of the differences obtained are added and compared to the sum of the applied thresholds. “Total_Threshold” is then computed in the first step by means of equation 3.12.

$$\mathit{Total\_Threshold} = \sum \mathit{partial\_thresholds} \tag{5.12}$$

After this operation (which is always positive in the first voting test), the result is stored and used as the new *Total_Threshold* value for further comparisons. Using this procedure, if a further region is observed to fit more effectively to the reference region (i.e. the result from the addition is smaller), uniqueness of the matching function is enforced and only this new region will be matched (the previous region will be left unmatched).

This matching methodology has been implemented as consecutive steps in a partial matching chain for every characteristic.

If some of the comparisons do not comply with the partial threshold, the inner loop is broken and reinitialized (there is no “else” statement), saving computational load.

As stated in the introduction of this section, several geometrical, sensor and segmentation assumptions have been considered, resulting in the following consequences:

- No geometrical correction is implemented. The two cameras are assumed to be parallel oriented and objects are far enough from the cameras. Then, only abscissa distortions are supposed to be perceived between both images.
- The depth map is approximated by parallel and non-overlapping planes.
- It has been assumed that every well-matched left centroid abscissa is higher than the right one (and they are equal when the region is located at infinity) and their

difference lower than 25% of the image range. This signifies that the matching regions are assumed to be close to each other and, thus, located far enough from the cameras.

- Both images are taken from the same camera height, so the scan-line to find matches can be assumed to be horizontal where only a range of +/- 25% will be tolerated when searching for matches.
- No region with an area below 0.1% of the image size will be catalogued as a significant region and as a result will not be matched.

As the images projected from each camera are different, several of the areas within the image might be projected in one of them, producing what is commonly referred to as the occlusion effect. These areas cannot be matched, as it has been widely discussed in stereovision literature [67]. It will be demonstrated that the method proposed in this setion leaves several regions where no matching occurs, and they will be indiscernible from occluded regions.

Let *(x,y)* be a descriptor of the centroid of a region in the left image, and *(x',y')* the same descriptor of the right image. Taking into account the geometrical assumptions detailed before, we can assume that:

$$(x,y) = (x'+d\cdot x\pm\varepsilon_x,y'\pm\varepsilon_x) \qquad (5.13)$$

where *d* is the distance (disparity) between the centroid abscissa, and *ε* the tolerance allowed in both directions.

Then, the main advantage of stereovision is the correspondence between differences in the *x* axis and the distance between the object and the cameras either once the cameras have been calibrated or the required assumptions have been made. The absolute distance of the centroid abscissas (in pixels) is measured for every matched pair of regions. Every pixel in each matched region will have a value of *d*. This gives a non-calibrated measure of the depth map.

In this algorithm, the left image is used as the background and used to compute the depth map. This is an arbitrary choice, without any loss in generality, based on the fact that the left image is the last one to be segmented and processed and, then, the footprint image (which is unique in the algorithm to save memory resources) corresponds to this one at the end of the segmentation process.

All the tested images encounter geometrical constrictions assumed by our algorithm, and where obtained from the Middlebury benchmark, with their truth depth map. As stated in the depth estimation section, the images that represent the computed depth maps are based on the left of the original images. The truth depth map is also provided over the left geometry of the image pairs [1].

First of all, it was run over the Tsukuba pair of images, which are presented in figure 5.29, the resolution of the image is 288×384.

Fig.. 5.29. Tsukuba gray scale pair of images: (a) left image and (b) right image.

In figure 5.30, the left version of the computed depth map based on the left image and the ground truth depth image of the Tsukuba pair are presented.

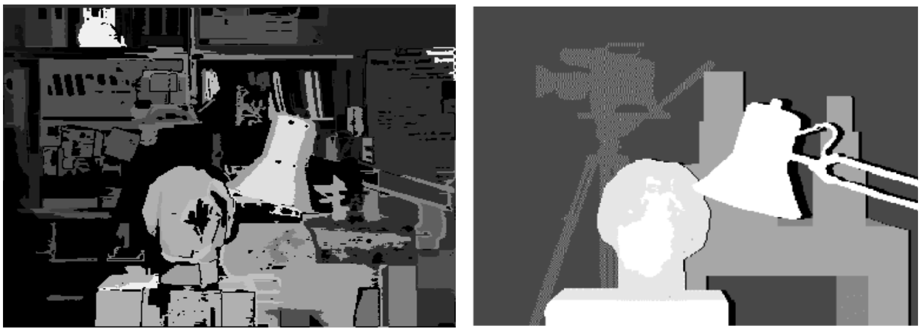


Fig. 5.30. (a) Tsukuba processed depth map. The image is normalized to its maximum value, which is achieved in the lamp. (b) Truth depth map.

These images have been segmented into 102 and 97 regions. The errors in non-occluded pixels, for a threshold of 2 (in absolute values) achieve the 55.9%. The error for all pixels is 56.6% and the error in discontinuities is 66%. These results will be discussed in the following section.

Regarding the computation time, the algorithm takes close to 24ms to the segmentation process of each image. As shown in [66], the segmentation time is quadratic related to the number of regions of the segmented images and directly related to the image size.

The characteristics vectors must be processed normalized and several of the descriptors are computed after the image segmentation from the extracted data. The time taken in this case is close to 90.5µs for each one. This is not significant regarding the segmentation and characteristic extraction processes. Finally, the matching has been carried out over the computed vector and not over the original images. In this case, the proposed algorithm takes up to 700µs to compare and match both vectors. We can see the total time consumption is around 50 ms (20fps) for this stereo pair.

We have carried out other tests over high textured images, using the Teddy and Venus gray scale images, which are shown here along (figure 5.31).

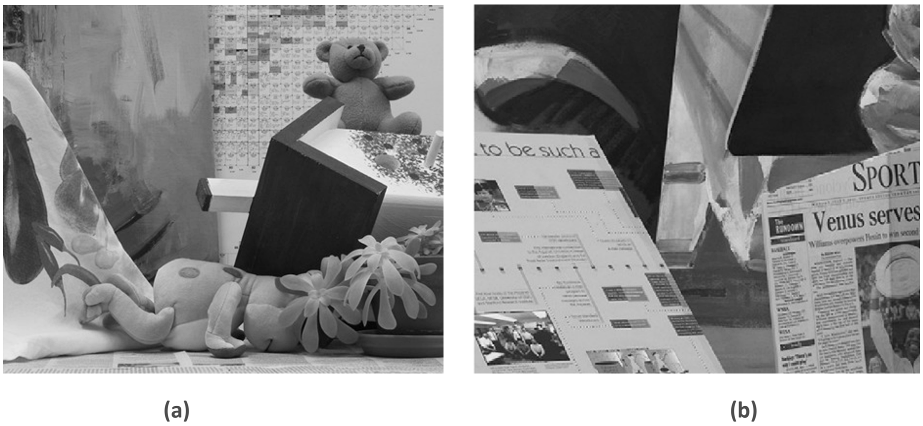


Fig. 5.31. (a) Teddy left image of size 375×450 and (b) Venus left image with the same size.

The corresponding results are presented in figure 5.32.

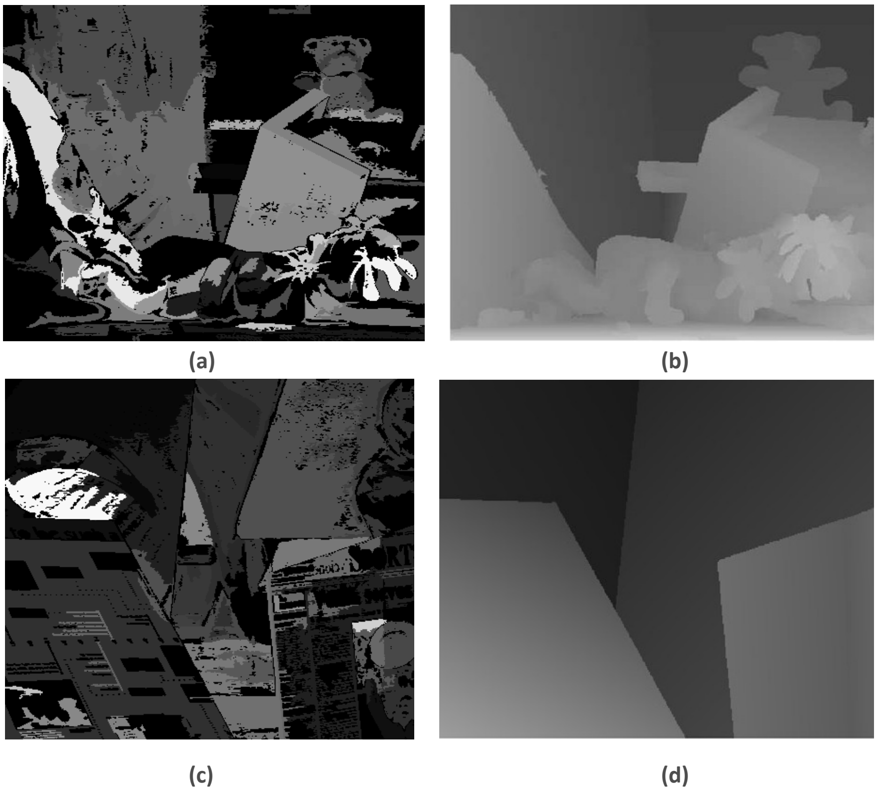


Fig. 5.32. (a) Teddy computed depth map, (b) Teddy true depth map, (c) Venus computed depth map and (d) Venus computed depth maps.

In figure 5.32, the computed depth map can be seen for the images presented in figure 5.31, with the corresponding truth depth maps.

	Non-occluded errors	All pixels errors	Discontinuities errors	Time (fps)
Teddy	79%	80.7%	68.7%	78.9ms (12.7fps)
Venus	73.9%	74.2%	68.7%	76.6ms (13fps)

Table 5.4. Errors and time performance in the Teddy and Venus pair of images.

As it has been discussed previously in this chapter, the matching process satisfies several of the different problems that arise in methods ranging from stereovision to image retrieval. The main contribution of the proposed algorithm is to solve the matching problem by comparing characteristics of regions instead of the regions themselves, reducing the computational cost, paying the price of higher error rates. Since images are segmented into no more than 255 regions, the computational efficiency increases with the size of the image as opposed to both windowed-areas and visual cue methods.

When designing the comparison steps, some decisions were taken, to set the partial thresholds. The value of the threshold limits the searching process, however this is not critical. The most significant parameter is the result from the voting process, this is automatically minimized during the vector comparison and, thus, matching is optimized among all the different matching possibilities.

The aim of this proposal is to retrieve depth map estimations with an important increase of the time performance, regarding other algorithms found in literature.

The truncation preprocessing has been presented apart from the algorithm, just for a better comprehension goal, but it is implemented on-the-flight in the region growing algorithm, thus, another advantage of the proposed algorithm is based on the fact that part of the preprocessing is carried out during the processing. No additional loops are then required in the program.

Moreover, no differential filter is required (as in [26, 54, 68]) for feature extraction and, in contra to these procedures, every segmented and recognized pixel is linked to a region, so the final construction of the depth map consists of searching within a Look-Up-Table (where the abscissa-differences are stored for each pixel of the matched regions) the depth value of the pixel is related to the identifier of the footprint image, and stored this value in the depth map image.

In contrast to other well known methods which perform image matching [2], we have replaced the task of comparing moving windows in both images by comparing two vectors that contain close to 100 terms in the Tsukuba images. The comparative process, due to its nested structure, allows the time spent on many operations to be saved when certain tests have not been passed. However, we can obtain unexpected results when analyzing the higher time required for computing the matching of smaller images or lower number of segmented regions. The explanation of this possibility is based on the number of steps required to be carried out in the nested comparing structure. If region descriptors are similar in values, the comparing structure needs to go through several steps, thus generating an increased computational load even when the number of descriptors is lower than that taken as reference.

The results obtained for the computed depth maps perceptually correspond to the truth depth map. However, errors are still evident: It can be seen that the gray scale segmentation based on brightness remains highly sensitive to noise. Other errors arise due to problems of matching small or undifferentiated figures such as the tins located behind the lamp in figure 3.30a. When looking at quantitative results, we see high error rates. These errors are due to the following reason. The disparity is computed from the centroids differences. But such centroids are not always in the correct place, since some regions can “overflow” the physical region, including some extra pixels. This fact offsets the centroid abscissa and, hence, the final disparity is biased.

Another relevant point to be made is related to the unmatched and non-segmented regions within the depth map (drawn in black). The majority of this lack of information is due to the minimum area condition required to segment a region. Most of the black regions are then not segmented (i.e. when the area is smaller than the minimum allowed) and, hence, not matched. The errors then propagate from this discrimination to the final depth map. A critical case of this error is presented in figure 5.32c, where the left written panel has not been matched. These sets of images are rich in color and textures, however their truncation into black and white produces a deficient segmentation, thus, deficient matching.

One of the main shortcomings to the proposed method is based on the segmentation quality. This particular area has been thoroughly investigated (see, for example, [69]) where adaptive region growing algorithms have been proposed to perform image segmentation similarly to the segmentation process carried out by human vision. The problems that arise with this technique are again based on the computational load required by the complex algorithms and the dependence on the image structure. In the compared algorithms, obtained results are very accurate, paying the price of a high complexity and, hence, computational load. The scope of the research work presented in this section has enforced the implementation of a context independent and faster algorithm resulting in a higher matching error rate. Thus, a compromise must be made between the segmentation, matching quality and computational complexity.

The main disadvantage of this algorithm has been shown to be the dependence of the matching quality in terms of the region growing quality. This dependence has been shown in figure 5.32 and its results, where deficient information (because of the gray scale preprocessing) is given to the region growing procedure and, hence, the matching is not satisfactory. If this approximation is worth enough, it will depend of the final application.

5.4.1.3. Single Channel Pseudo-Color Segment-based Stereo Matching Algorithm

The previous algorithm presents some problems, regarding to the loss of information because of gray scale truncation. On the other hand, color processing is, normally, three times more complex, because of the three channels that every color representation of an image has for each pixel.

The field of research referred to as color image processing has been widely studied over the past few decades, and this is partly due to the fact that it is closely related to the process of human vision [12, 70, 71].

The extra information contained within the color space when compared to gray scale images is widely known [72], and can clearly be seen in figure 5.33.

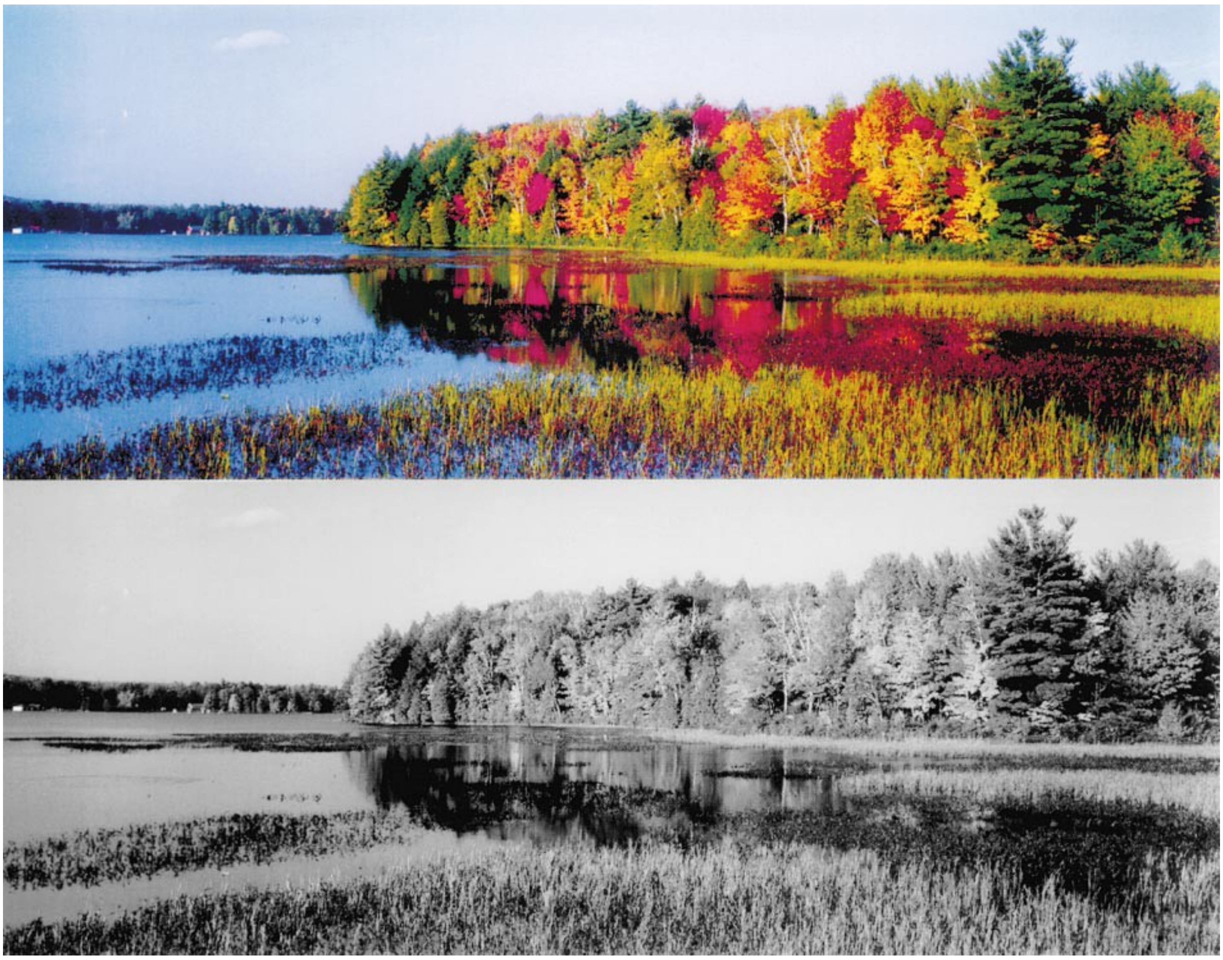


Fig. 5.33. Advantages of color vision [72].

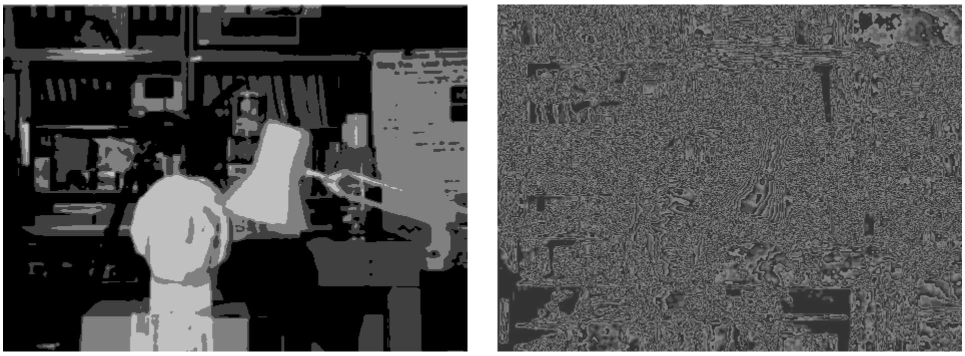
Color is commonly presented as the combination of three components, Red-Green-Blue (RGB). Other possibilities come from transpositions of this space into three other coordinates, such as YIQ or YUV (for NTSC and PAL composite video), HSL, HSV, etc. No matter which color space is used, there are always three components, although some of them may be compressed as in the case of composite video.

The basic principles of the improved and novel algorithm proposed in this section are listed as follows:

- Not every bit of a pixel carries the same information: most relevant bits have more information than the least bits.
- During the process of image segmentation it is only important to compare relevant information.
- A pseudo-color image (PCI from hereinafter) can be built from a color image and maintains the majority of the advantages associated with both color and gray scale segmentation.

This PCI is computed from the three channel color image, converting the relevant information to a gray scale. This reduces the complexity of the segmentation process, while at the same time maintains the most relevant color features.

A standard color image requires 3 bytes per pixel. However, not every byte is equally relevant for the interpretation or description of the image. For example, the first two bits contain much more information required for the segmentation algorithm than the last two bits; this is illustrated in the following image where only the red channel has been considered. From this figure we can appreciate the lack of information provided by the last two bits of the red channel.



(a) (b)

Fig. 5.34. Two most (a) and least (b) significant bits of the red channel acquired from the Tsukuba right image.

One option which has been considered during the development of the algorithm has been to reduce the computational load by only considering the first two bits of each image channel, creating a gray scale image where each pixel is composed of only one byte, as illustrated in table 5.5:

B_8	B_7	G_8	G_7	R_8	R_7	0	0

Table 5.5. Pseudo-color byte structure.

In Table 1, ‘B_8’ and ‘B_7’ represent the two most significant bits of the *Blue* channel value of the pixel, and ‘*G*’ and ‘*R*’ represent the *Green* and *Red* channel values of the same pixel, respectively.

However, this particular solution presents two significant problems which must be dealt with:

- High sensitivity to noise in gray parts of the image. When the value of the gray level of the pixel is close to the two most relevant bits in each channel, as a result false colors appear.
- Dark areas are set to black. Since only two bits are taken into account, levels under 25 % (in 8-bit gray scale images) are set to zero.

These two problems have been solved by implementing color clustering. This involves identifying the dominant colors in an image, and setting the rest of the bits with information regarding relevant color levels. This process can be carried out using the following steps: RGB conversion to *Hue*, *Saturation* and *Value* (HSV) color space, color clustering with the *hue* component and calculation of the pseudo-color gray scale image.

The main advantage of the HSV color space is that it provides decoupled chromatic information from the intensity and saturation. This transformation results in much improved color clustering that can then be implemented within this color space.

The color space conversion has been carried out by using the following transformations [73]:

$$
\begin{aligned}
&\max = Max(R,G,B); \min = Min(R,G,B);\\
&H = \begin{cases} (G-B)/(\max-\min), (R=\max)\\ 2+(B-R)/(\max-\min), (G=\max)\\ 4+(R-G)/(\max-\min), (B=\max) \end{cases}\\
&V = \max;\ H = H*60; If(H<0) H = H+360\\
&S = \begin{cases} 0, If(\max=0)\\ 1-\min/\max, Otherwise \end{cases}
\end{aligned}
\tag{5.14}
$$

The preprocessing principle consists in clustering colors in the image to a reduced number of them, assigning homogeneous values to pixels when their color is close enough. This transformation produces an image with few colors, which can be expressed with fewer bits and, finally, compressed to gray scale. We implement the clustering over the *hue* component because is the one which keeps the chrominance information of each pixel.

The histogram of *hue* component is computed for an image (assuming that both images have similar chromatic components) and is then processed by the algorithm which runs over the histogram and obtains the main components:

The maximum number of colors (*MAX_NUM_OF_COLORS*) is a constant that forces the threshold increase of the local minima variation to fit the palette of colors to the desired one. The *newHistogram* array stores the same *H* value until the algorithm finds a local minima or an isolated group of colors above another threshold, when *counts* increases its value representing a new color.

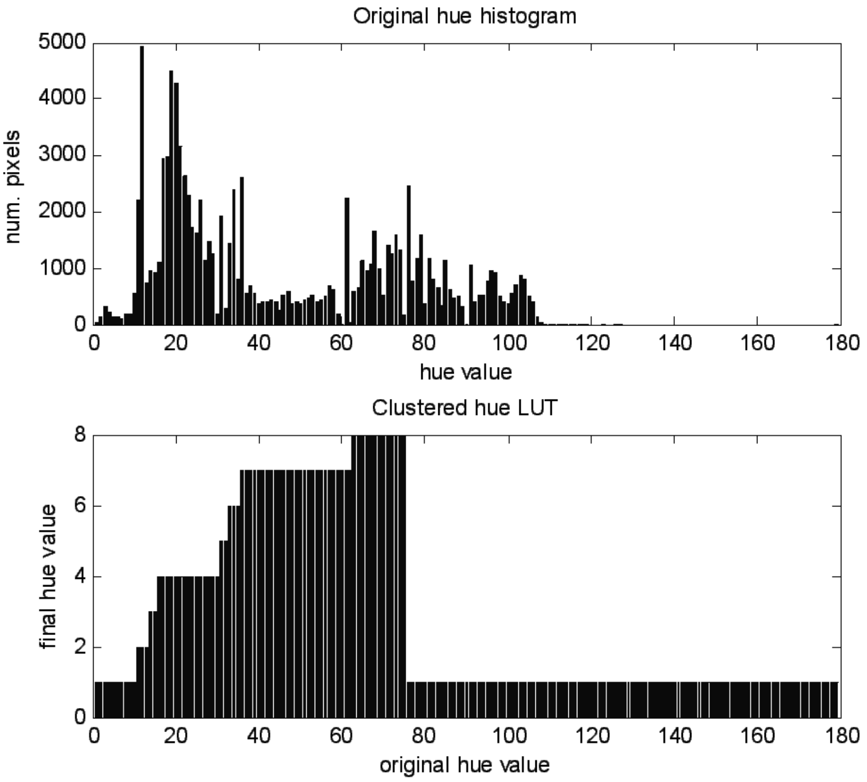


Fig. 5.35: (up) Original histogram of Tsukuba right image (ranged [1:180], horizontal axis) and (down) Clustered Look-Up-Table to 8 colors (ranged [1,8], vertical axis).

As shown in figure 5.35, the array obtained may not be considered as a proper histogram, but as a 1D-LUT where a value (ranging between 1 and 8) is assigned to every Hue value of each pixel in the original image.

In order to avoid the near-gray levels that produce false color effects when truncating less significant bits, the *Saturation* and *Value* components must be verified to be higher than a threshold for each pixel. This means that some originally gray pixels will be processed in a gray scale format. The complete transformation for each pixel is as follows:

```
if O(Si,j) > SAT_THRESHOLD & O(Vi,j) > VAL_THRESHOLD Then
        F(Hi,j) = O(Hi,j);
        F(Si,j) = 255;
        F(Vi,j) = O(Vi,j)&MASK;
else
        F(Hi,j) = O(Hi,j);
        F(Si,j) = 0;
        F(Vi,j) = O(Vi,j)&MASK;
end if;
```

In this pseudo code, the terms *O(…)* and *F(…)* are the original and final images, respectively. $H_{i,j}$, $S_{i,j}$ and $V_{i,j}$ are the *Hue*, *Saturation* and *Value* component , respectively, of the *(i-th,j-th)* pixel and *MASK* is the constant that maintains the two MSBs rejecting the rest ones.

The result of the aforementioned process is the conversion of the original image to a scalar image. Figure 5.36 illustrates two representations of the final PCI image (false colors have been introduced to aid visual perception):

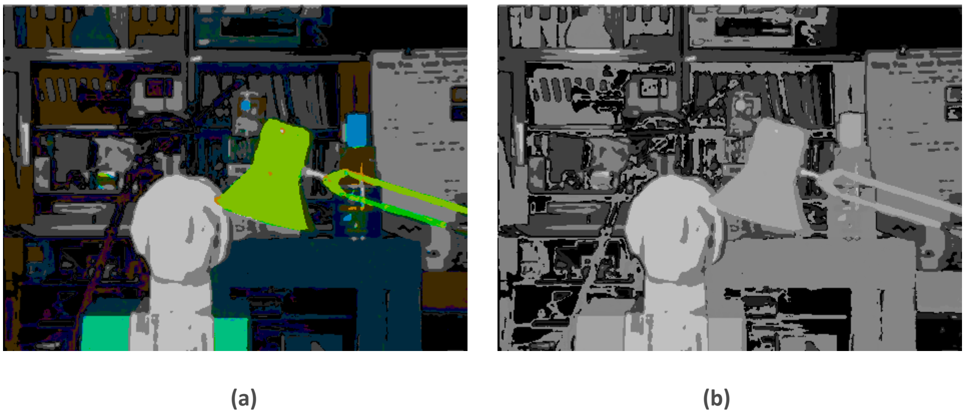

(a) (b)

Fig. 5.36. (a) Tsukuba right image after color clustering (1 byte/pixel). False color representation. (b) PCI shown in a gray scale image.

Since the level of *Saturation* has been reduced to one bit (whether or not it is completely saturated), the color palette is forced to be represented by 8 colors. As effect of the minimum saturation constraint, some pixels are left in gray scale.

The *Value* of every pixel is truncated to its two MSB's, the complete pixel information can thus be stored in 6 bits, which maintain the most relevant information of each pixel. As a result, a gray-scale image can be built, which is then segmented as a non-ambiguous gray scale image. The final PCI gray-scale image is presented in figure 5.36b. In these images, each byte has the following structure:

S	C_1	C_2	C_3	V_8	V_7	0	0

Table 5.6. Clustered pseudo color byte structure.

In this table, *S* is the saturation bit, C_x the descriptor of the dominant color (allowing 8 different colors) and V_8 and V_7 the two MSB of the *Value* component.

To verify the reliability of this procedure, it has been applied to the Tsukuba pair of color images shown in figure 5.37, which have a resolution of 384×288 pixels.

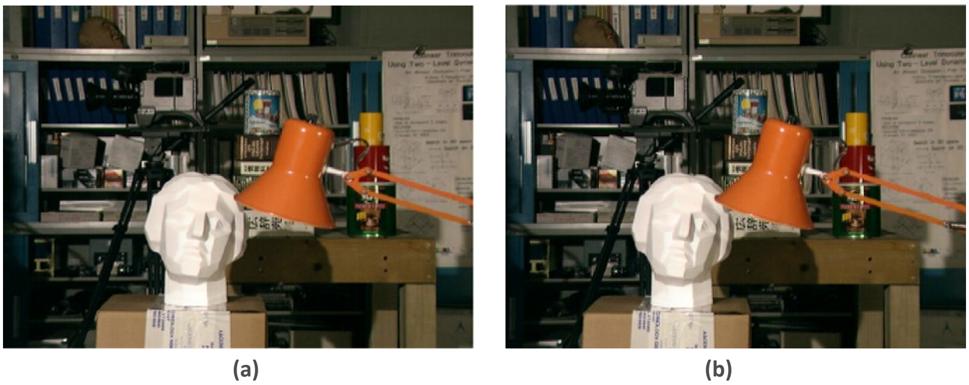
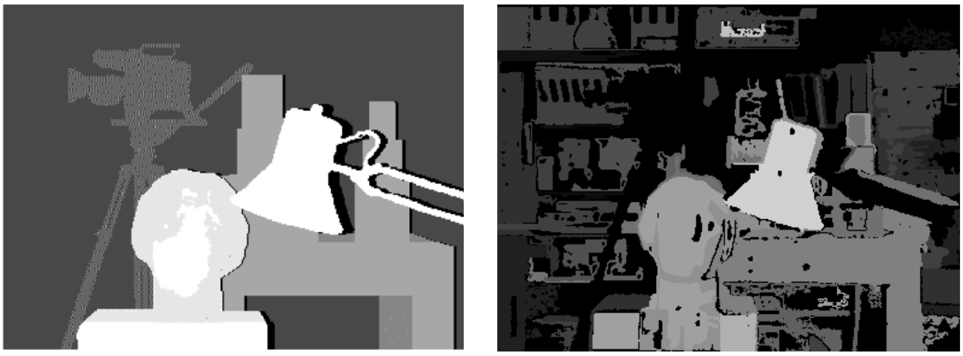

(a) (b)

Fig. 5.37. Tsukuba color images pair.

Figure 5.38 shows the real and computed depth map acquired using the proposed procedure.



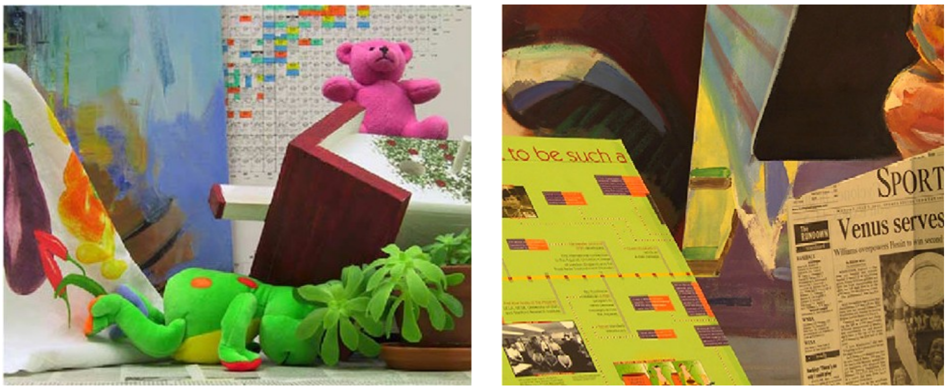
(a) (b)

Fig. 5.38. (a) Real depth of Tsukuba images. In black, the occluded parts. (b) The computed depth map. In black, unmatched regions and occluded parts.

The time required by the algorithm to obtain this result is 77.4ms (12fps), achieving a real-time performance.

The quantitative results, obtained by means of the Middelbury web page [1], show that the error in non-occluded pixels achieves the 46.9% of pixels for a threshold of 2.

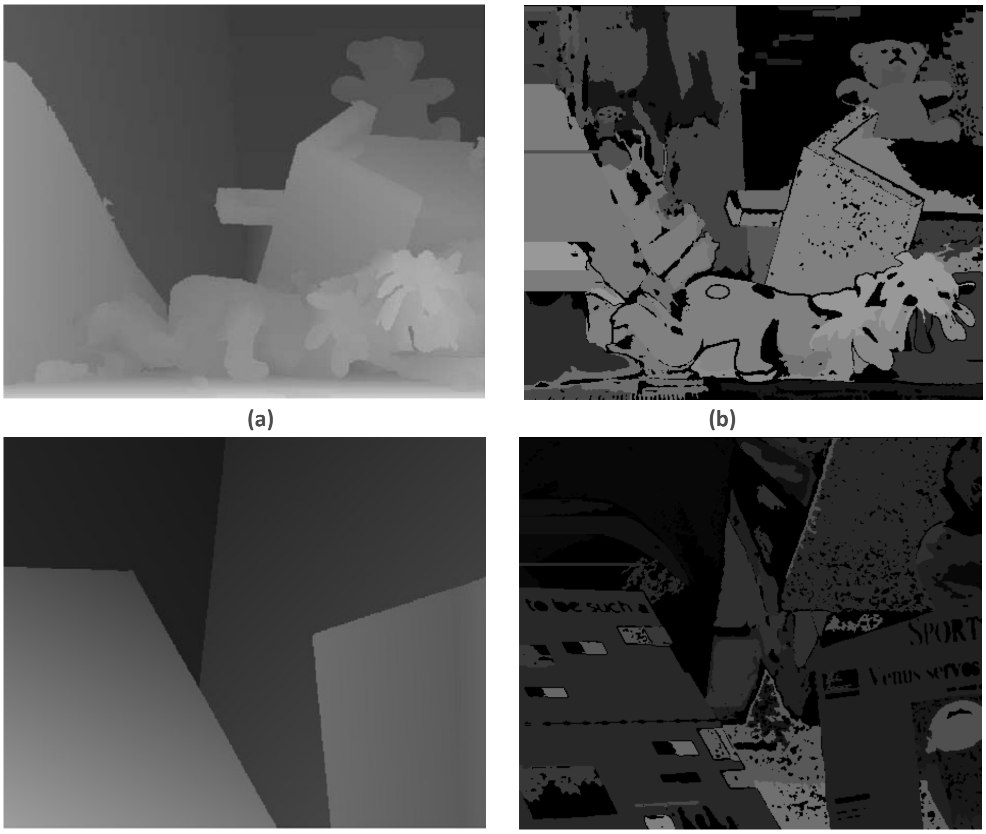
Further tests on different color standard image pairs have also been carried out. The set of figures where the proposed algorithm has been tested are shown in figure 5.39, along with their corresponding true depth maps. The Teddy set has a resolution of 450×375 pixels, while the Venus set has 434×383 pixels.



(a) (b)

Fig. 5.39. (a) Teddy and (b) Venus images.

Image processing results on each image pair is presented in the figure 5.40, close to the true depth maps.



(a) (b)

(c) (d)

Fig. 5.40. (a) Teddy true and (b) computed depth map. (c) Venus true and (d) computed depth map.

In these cases, the time required to compute the depth maps is 114.2ms (8fps) for the Teddy image, and 111.8ms (8fps) for the Venus image.

The quantitative evaluation was performed in the same way of the Tsukuba image pair. In these cases, the error in non-occluded pixels for the Teddy and Venus image pairs is 60 and 77.2%.

The main goal of the preprocessing and color clustering algorithm was achieved; it may be observed that the preprocessing has assigned each pixel with a value which ranges within a lower number of levels and that automatically creates refined regions, these will then be segmented using a comparative analysis of their respective values. It is important to point out here that the levels obtained are not related to the human vision system. Also, the order of the bits shown in table 5.6 is irrelevant, since an exact comparison is done in every step of the segmentation algorithm.

It is interesting to perform a qualitative comparative analysis between the proposed algorithm for color images and the previously developed algorithm applied to gray scale images.

In figure 5.30a, 5.32a and 5.32c critical errors may be observed which are due to the gray scale segmentation and post matching process. These errors result from the lack of information of the gray scale image. By using the color based algorithm presented in this proposal most of these errors have been corrected.

Detailed differences in the Tsukuba image were not very important. However for the case of the lamp several differences may be observed where noise components are seen to be removed. This may also be observed in the tins behind the lamp, which have not been matched or segmented in the gray scale version. In these images, the color based algorithm provides improved segmentation and matching results, however several new errors arise. Over the left panel in figure 5.39b, the dark part of the image has been segmented into several different regions, where some of these result in a false match and are given as bright areas in the depth map (figure 5.40d).

All remaining results are seen to improve when using the novel color-based algorithm, where a more detailed depth map has been obtained. However several errors are still present which are related to the segmentation process. An example of such errors may be observed in the different areas of the left panel presented in figure 5.39b which provoke the errors shown in figure 5.40b. These can be explained by the nature of the segmentation algorithm. As the areas with the same value (regardless of the image being in gray or color scales) are processed as being the same region, any depth differences within the same region (i.e. left panel of image shown in figure 5.39b) have not been computed, and only a mean value is provided. This particular effect may also be appreciated when observing the floor of figure 5.40b. Additionally, some differentiated areas inside this panel (color blocks, for example) have been processed as different regions, thus, the depth has been computed separately. This effect produces the result that can be observed in the image presented in figure 5.40d.

5.4.2. Points-based Fast Depth Map Estimation

The main idea implemented in the dynamic programming strategy was scanning and matching pixels in one dimension, but we found some troubles regarding inter-scan lines streak noise. To overcome this problem, a fast proposal can be detecting only vertical edges to stabilize vertical coherence.

Moreover, taking into account the idea firstly proposed in [74], we can arrive to the following geometrical assumptions:

- Edges in the epipolar line have no depth information. The depth information is given by the following expression:

$$depth_i = f(\vec{e} \times \overline{u_i}) \tag{5.15}$$

where ***e*** is the director vector of the epipolar line and ***u****ᵢ* the director vector of an edge in the image. The depth is a function of the cross product because it also depends of other scene features. Thus, the depth information is contained in the normal projection of any edge to the epipolar line when cameras are placed parallel.

- We assume that the scan lines for point extractions and matching are horizontal. Since the *fronto-parallel hypothesis* is taken into account, and the cameras setup fits with this constraint, we can assume that every physical (and not occluded) point in one image must be found at the same height in the other image. The horizontal displacement will be used to compute the depth.

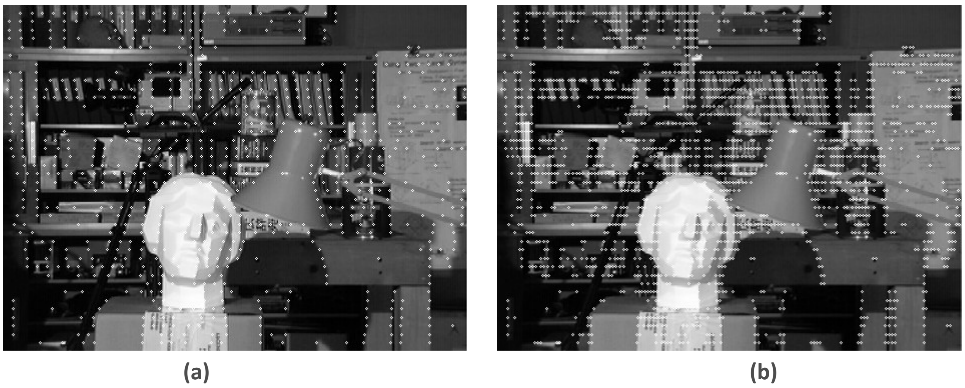
In this scenario, we propose the following solution to the depth map estimation:

- The depth estimation can be solved with an approximation, aiming to process only the most relevant information in the problem, avoiding extra and costly calculation.
- As a consequence of the first geometrical assumption, only vertical edges are relevant. It allows performing a horizontal 1D differential filtering to extract the relevant points, instead of using 2D masks.
- We can compute only vertically relevant points over a low number of lines across the image, with the minimum number of operations.
- Implement a point matching based of some point features, assigning to each point a depth value.
- Interpolate the value between matched points.

The proposed algorithm works over gray scale images.

The first task to be implemented is the relevant points extraction. This is done with a horizontal differential mask of size 1×5, with the following structure: {-1,-1, 0, 1, 1}. The horizontal size of the window allows reducing the effect of the pepper and salt noise as well as the Gaussian blur. The convolution over this mask is compared to a threshold for some pixels in the scan line.

An extra assumption is considered: depth relevant points are not close to other relevant points. This assumption is equivalent to say that variations in the depth of the image are not very close one to each others. This is implemented in the algorithm forcing a jump of “n” pixels in the scanning task when a relevant point is found. For illustration purpose, two images of this processing over the left Tsukuba image are shown in figure 5.41 for two different values of this constraint.



(a) (b)

Fig. 5.41. Vertical relevant points extracted forcing a separation of (a) 9 and (b) 2 pixels. In both cases, there are 40 scan lines.

For every found point, the following characteristics are stored:

- Sum of the previous pixels, given by the convolution of the first two elements of the mask.
- Sum of the next pixels, given by the convolution of the last two elements of the mask.
- Coordinates of the pixel.

- Sum of the 3 upper and down pixels to obtain better descriptors and, hence, matching. This 2D processing in done only once a pixel is found as relevant, avoiding a 2D convolution for most of the pixels of the image.

Once the most relevant pixels have been extracted, with the proposed descriptors, we can implement a hierarchical and nested conditional comparison chain to find the optimum match for each point. In each iteration, the minimum Sum of Absolute Differences (SAD) is stored with their corresponding pixels. After scanning the whole line, the best match combination is retrieved and its corresponding pixels matched.

From this process, we get the best matched points and, with their abscissa descriptor, we can estimate their depth. Notice that this matching structure runs over a limited number of elements, much fewer than the pixels contained in the original image.

We find a problem, once the relevant points are matched: the interpolation. Figure 5.42 shows the top view of some matched points and some possibilities of interpolation.

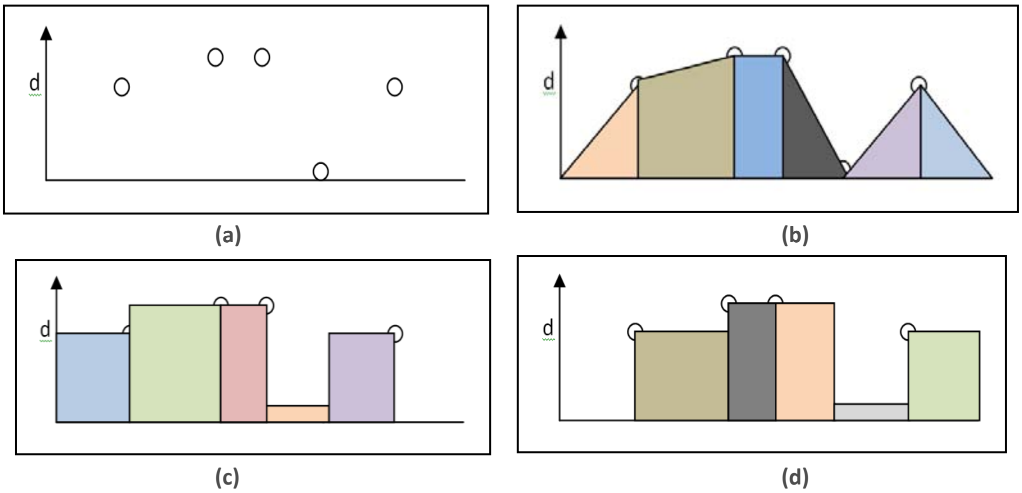


Fig. 5.42. Top view of (a) Matched points at distance "d". (b-d) Polynomial, right guided and left guided interpolation respectively given (a) points.

There is, hence, an ambiguity when interpolation process is needed. I will present results with *left guided* interpolation.

Since not every row of the image is scanned, the algorithm deals with "images" of L×M for images of N×M pixels, being L the number of scan lines. Thus, the final image is a stretched version of the estimated depth map and must be resized to its original size N×M. This allows managing a simplified version of the images, where any other processing will be much lighter than in the original size image.

The most important post process implemented over the stretched image in our proposal is a vertical median filter of window size 3×1, processing the estimated stretched depth map before resizing it.

The resulting matching points of Tsukuba left image is shown in figure 5.43.

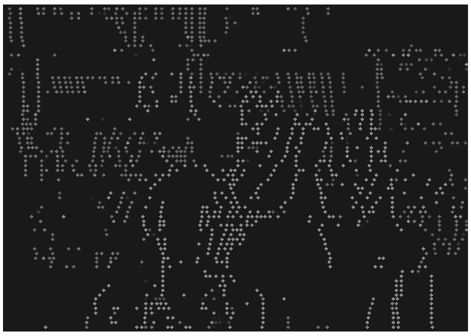
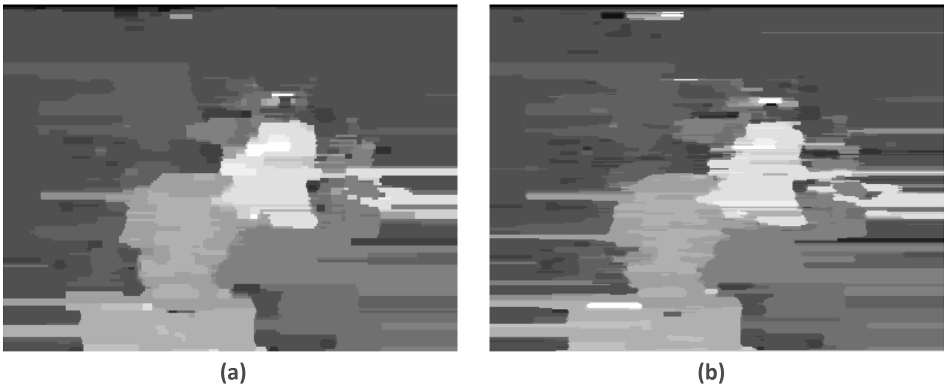


Fig. 5.43. 80 scan lines point matching from figure 5.41.a.

This extraction and matching procedure is very fast, and only takes the half of the time of the complete depth map estimation, around 4ms. Thus, interpolating and painting the whole depth map appears to be a high cost process.

Taking the Tsukuba pair of images (288×384), we get the depth maps estimations for the half and full resolution scan lines shown in figure 5.44.



(a) (b)

Fig. 5.44. Tsukuba depth map estimation with (a) 144 and (b) 288 scan lines.

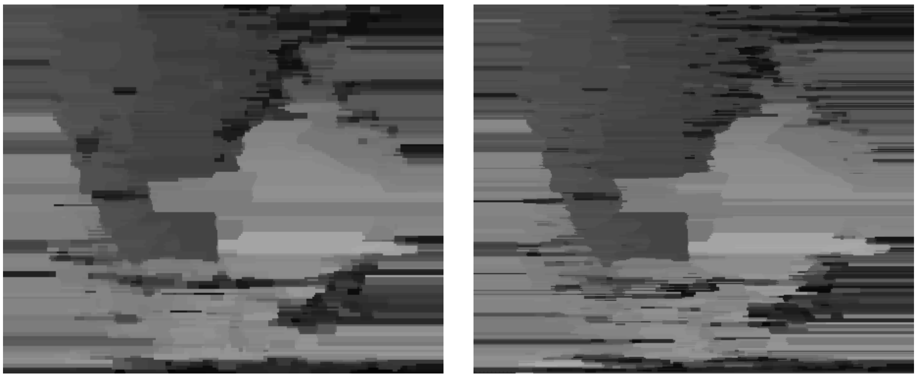
As explained before, results sacrifice accuracy to achieve high processing speed:

Number of scan lines	Time (ms)	fps	Non-occluded Errors(%)
(a) 144	11.9	84	10.5
(b) 288	23.2	43	11.3

Table 5.7. Performance of the algorithm for different number of scan lines in the Tsukuba pair of images. Threshold for the error estimation: 2.

The time delayed by the algorithm shown in table 3.6 considers the scanning, the matching and the post processing median algorithm.

Figure 5.45 shows the results for the Teddy pair of images. The size of each image is 375×450 pixels and the median filter size used in the post processing is 5×1.



(a) (b)

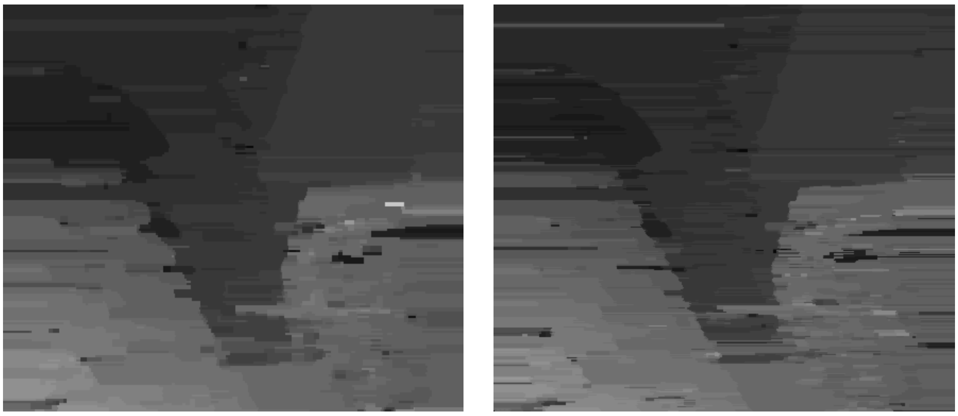
Fig. 5.45. (a) Teddy depth map estimation with 187 and (b) 375 scan lines.

The time analysis is shown in table 5.8.

Number of scan lines	Time (ms)	fps	Non-occluded Errors(%)
(a) 187	20	50	45.4
(b) 375	41.6	24	45.4

Table 5.8. Performance of the algorithm for different number of scan lines in the Teddy pair of images.

Finally, the Venus pair of images are processed and their depth maps presented in figure 5.46 and table 5.9. These images have a resolution of 383×434 pixels.



(a) (b)

Fig. 5.46. (a) Venus depth map estimation with 191 and (b) 383 scan lines.

The time analysis is shown in table 5.9.

Number of scan lines	Time (ms)	Fps	Non-occluded Errors(%)
(a) 191	17.5	57	15.2
(b) 383	35.7	28	14.9

Table 5.9. Performance of the algorithm for different number of scan lines in the Venus pair of images.

Nowadays there is no algorithm that achieves accurate depth maps estimations without investing a huge amount of computational resources and time.

The quality of depth maps estimations presented in [38], [56] and [57] are higher than that proposed in this section, but regarding the time performance, we find interesting results. Hirschmüller *et al.* obtain this result over a 450MHz processor at 4.7fps (notice that the image size is 240×320 in this experiment) [38]. Hong and Chen get their results with a 2.4GHz PC after 3 seconds (image size 288×384). In the case of Klaus *et al.*, the computation time required is higher than 14 seconds in a 2.21GHz machine [57] (the size is not specified in this work). We can easily see the improvement in terms of performance of our algorithm, since our results for the same image achieve a frame rate of 84fps in slower processors and best case (except the case of Hirschmüller *et al.*).

Among the results found with point matching algorithms, we can highlight the following ones: Result shown in figure 5.21a is accurate, but the algorithm with which it was extracted has a high computational load, since it is a multi scale approach with 2D derivatives to find control points. The author gives no estimations about complexity, order or time in this case. Neither they do in the case of figure 5.21b, but we can guess that resources and time consumption are quite high since they use 2D spatial derivatives to find the relevant points.

Interesting results, because of their accuracy, were given in figure 5.21c and 5.21d. Let's remember that those results were extracted in a Pentium IV (@2.4GHz) delaying 11.1 seconds and 4.4 seconds for the Venus and the Tsukuba pairs respectively.

Regarding to time consumption, the solutions proposed in the literature yields to results far from those found and shown in this proposal. In [75], the point extraction and the matching process last between 99 and 472 seconds. Other works show processing times for points extraction and matching from 5 seconds [76] to more 20 seconds (and easily reaching hundreds of them) [77], depending on the implementation.

Regarding to the complexity of the algorithm, we can appreciate linearity in the computation time as a function of the number of lines. Moreover, the order of the algorithm we present and analyze in this section is $O(M \cdot L)$, being M the number of columns of the image and L the number of lines to be scanned. No dependency of the height of the image was found. Just for comparison purposes, we can take the example given in [78]. The algorithm proposed in this section matches points for fingerprint identification. It has a complexity of $O(n^2k^2\log n)$, being k the operator size and n the size of the image.

This essential difference is due to the highly specific approach adopted in this work: we don't deal with rotation or scale variant images, as is done in [78]. We can assume that the epipolar lines are horizontal and identical in both images and, hence, we don't have to search the feature points in a 2D space. This constraint reduces in one dimension the complexity of the problem. We can exploit all these specification of a specific stereo vision problem to adjust the searching and matching algorithm to make it as simple as possible. Then, its complexity falls down dramatically and although final results are not very accurate (it depends of the final application, as said), we find it interesting as a new way to solve the stereo matching problem.

The interpolation is the bottleneck of our proposal. Not only because of the time consumption (half of the total processing time), but also because of its results. Many errors can be appreciated in depth maps shown in figures 5.44-5.46. These errors are more due to problems in the interpolation process than mismatches, as it is illustrated in figure 5.43. In that figure, we can see the points corresponding to their real position (left image as reference) and with their corresponding depth (in gray scale).

5.4.3. Dense Fast Depth Map Estimation

Starting from the previously implemented algorithm, we can extend the idea of relevant point matching to a dense formulation, keeping some of the advantages of the previous algorithm in terms of computational load, but achieving more accurate results.

This approximation takes into account the same hypothesis than that of the previously presented algorithm, as done in the dynamic programming approximation.

The geometrical assumptions are those explained [27, 31]. We assume, as in the previous algorithm, that the scan lines for points extractions and matching are horizontal. Since the *fronto-parallel hypothesis* is taken into account, and the cameras setup fits with this constraint, we can assume that every physical (and not occluded) point in one image must be found at the same height in the other image. The horizontal displacement will be used to compute the depth. The next assumption to be made is that every neighbor of a pixel has similar disparity, what we called *local coherence constraint* [48]. Of course, this constraint is not true in edges and occluded pixels, but it allows us to perform a pixel-by-pixel matching assuming this constraint until a comparison threshold is surpassed.

In this scenario, I propose the following solution to the depth map estimation:

- Finding a first relevant point for each scanline, by means of a horizontal and unidimensional gradient operator.
- Finding the corresponding point in the other image with the same operation.
- Try to match pixel by pixel, defining an acceptance threshold and allowing some outliers (to reduce the effect of impulsive noise).

The proposed algorithm works over rectified gray scale images.

The first task to be implemented is the relevant points extraction. This is done with a horizontal differential mask of size 1×5, with the following structure: {-1,-1, 0, 1, 1}. The horizontal size of the window allows reducing the effect of the Gaussian blur. The convolution over this mask is compared to a threshold to decide whether it is a relevant starting point or not. Once two corresponding pixels are found in one line, the pixel-by-pixel match is implemented every scanned line.

We have, then, the possibility of scanning or not every line of the original pair of images. When not every row of the image is scanned, the algorithm deals with “images” of L×M for images of N×M pixels, being L the number of scan lines. Thus, the final image is a stretched version of the estimated depth map and must be resized to its original size N×M. This allows managing a simplified version of the images, where any other processing will be much lighter than in the original size image, as will be shown later.

We find three degrees of freedom in this algorithm:

- Percentage of horizontal lines scanned
- Threshold for pixel acceptance as non-outlier
- Number of allowed outliers

The number of scanned lines can vary, taking all of them or a subset. This possibility opens the door to a better optimization between accuracy and processing speed, finding an specific agreement between these two factors depending on the final application.

Regarding the threshold tolerance to accept or not a new pixel (the SAD between him and the previous one), is an important parameter in the algorithm speed, as it will be shown in the Results section. This parameter is, moreover, related to the number of accepted outliers before searching a new matching condition.

In the following section, we will present an independent optimization of each parameter.

The main post process implemented over the stretched image in our proposal is a vertical median filter of window size 3×3, processing the estimated stretched depth map before resizing it, when not every line has been examined.

Taking the Cones pair of images (450×375), we present the number of errors in non-occluded and discontinuity pixels in terms of the number of scanned horizontal lines. Under the number of %, it is written the frames per second of each implementation. Threshold is set to 7 and the number of allowed outliers to 4. The threshold to consider a pixel as wrong is 2.

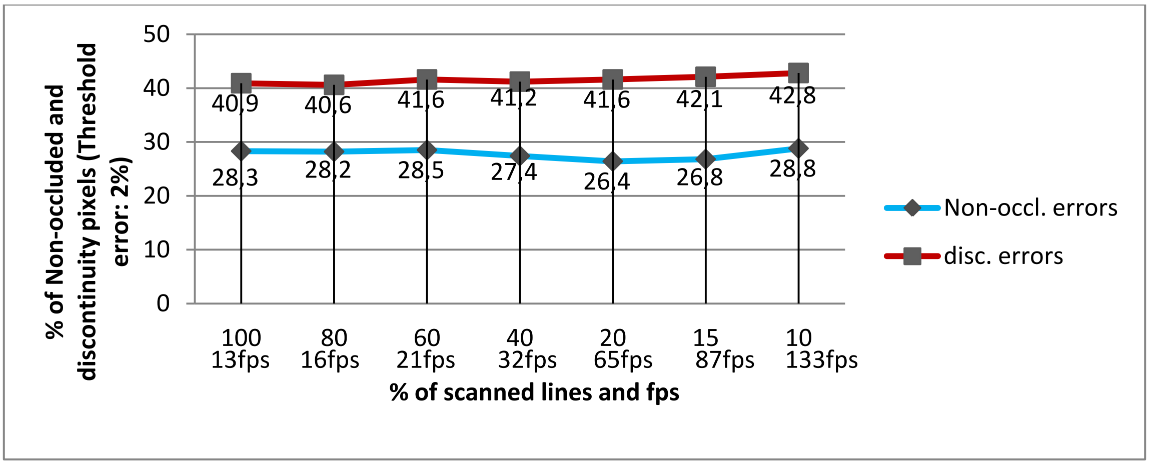


Fig. 5.47. Errors and time performance regarding the percentage of lines scanned.

We can easily see a minimum at 20% of scanned lines for the non-occluded pixels, while the discontinuity error pixels achieve a minimum at 80%. Both images are shown in figure 5.48.

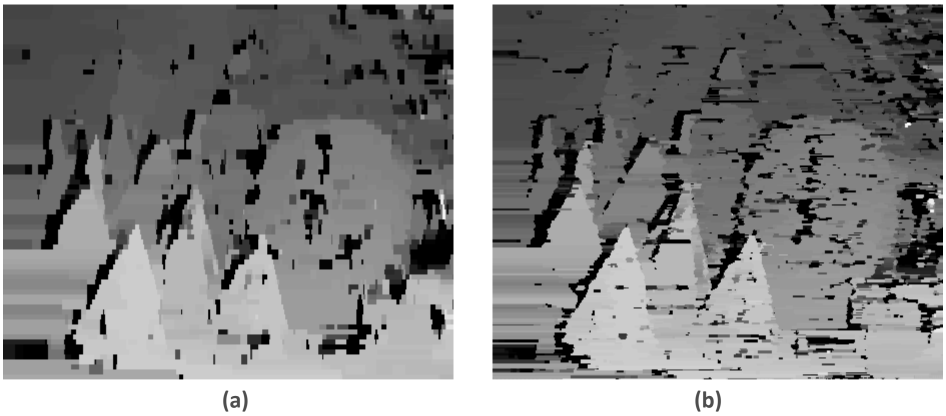

(a) (b)

Fig. 5.48. Cones depth map estimation for (a) 20% and (b) 80% of scanned lines.

In figure 5.49, we present the evolution of the non-occluded pixels error in terms of the threshold, measured in absolute value (the rank for the pixels value is [0,255]), with the time performance in frames per second. In this representation, the number of scanned lines is 100% and the number of outliers is 4.

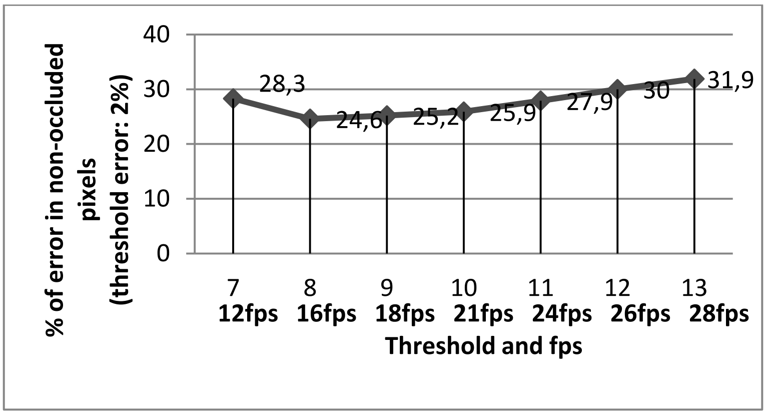


Fig. 5.49. Error of non-occluded pixels and time performance against the acceptance threshold.

Finally, we show in the following graph the evolution of non-occluded pixels error and time performance against the number of allowed outliers, with threshold 8 and 100% of lines scanned.

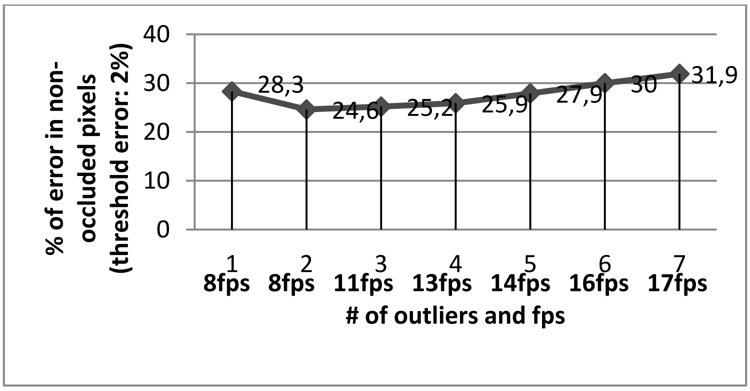
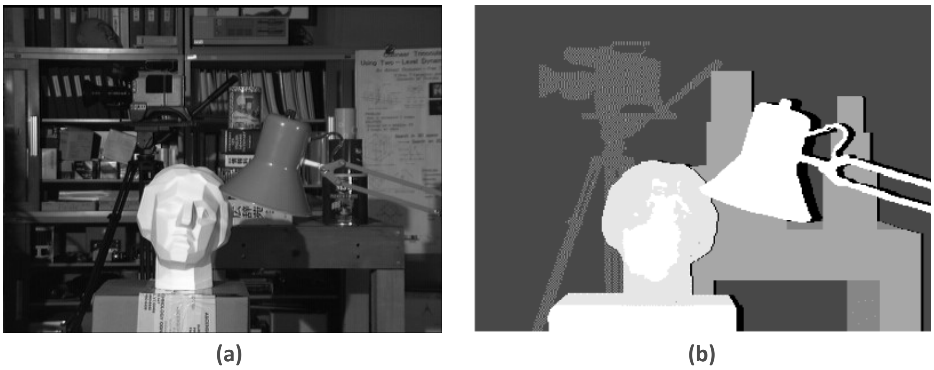


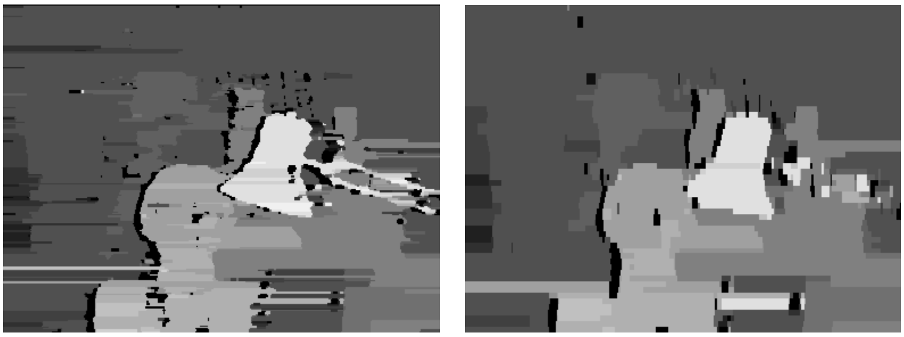
Fig. 5.50. Error in non-occluded pixels and time performance in terms of the number of allowed outliers.

When processing the Tsukuba pair of images, we find the following results for 80% and for 20% of scanned lines, threshold of 8 and 3 allowed outliers.



(a)

(b)



(c) (d)

Fig. 5.51. (a) original Tsukuba left image, (b) ground truth, and depth map estimation with (c) 80% and (d) 20% scan lines.

As explained before, results sacrifice accuracy to achieve high processing speed as shown in table 5.10.

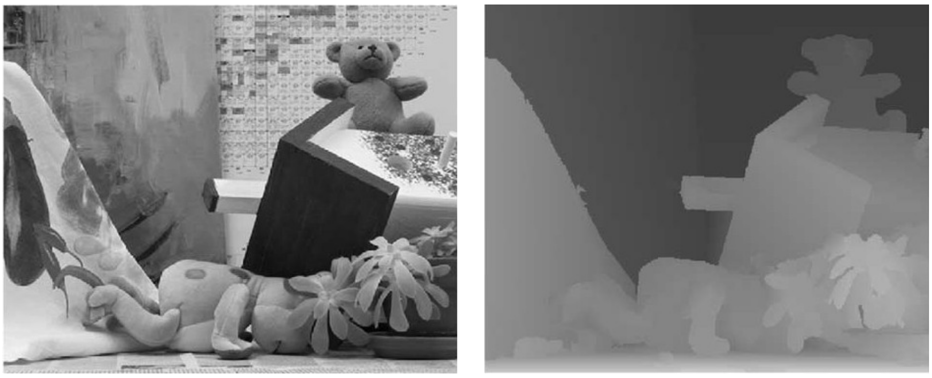
Number of scan lines	Time (ms)	fps	Non-occluded Errors(%)
(a) 80%	11.3	88	8,99
(b) 20%	2.9	347	9.98

Table 5.10. Performance of the algorithm for different number of scan lines in the Tsukuba pair of images. Threshold for the error estimation: 2.

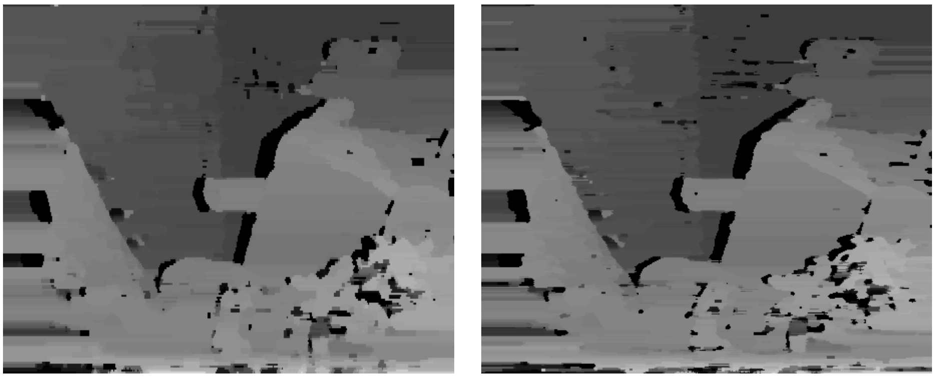
The time delayed by the algorithm shown in table 5.9 considers the scanning, the matching and the post processing median algorithm with 3×3 window.

As usually happens in dynamic programming approach, “streak” lines appear, which cannot completely be removed by the median vertical filter.

Figure 5.52 shows the results for the Teddy pair of images. The size of each image is 375×450 pixels and the median filter size used in the post processing is 5×5.



(a) (b)



(c) (d)

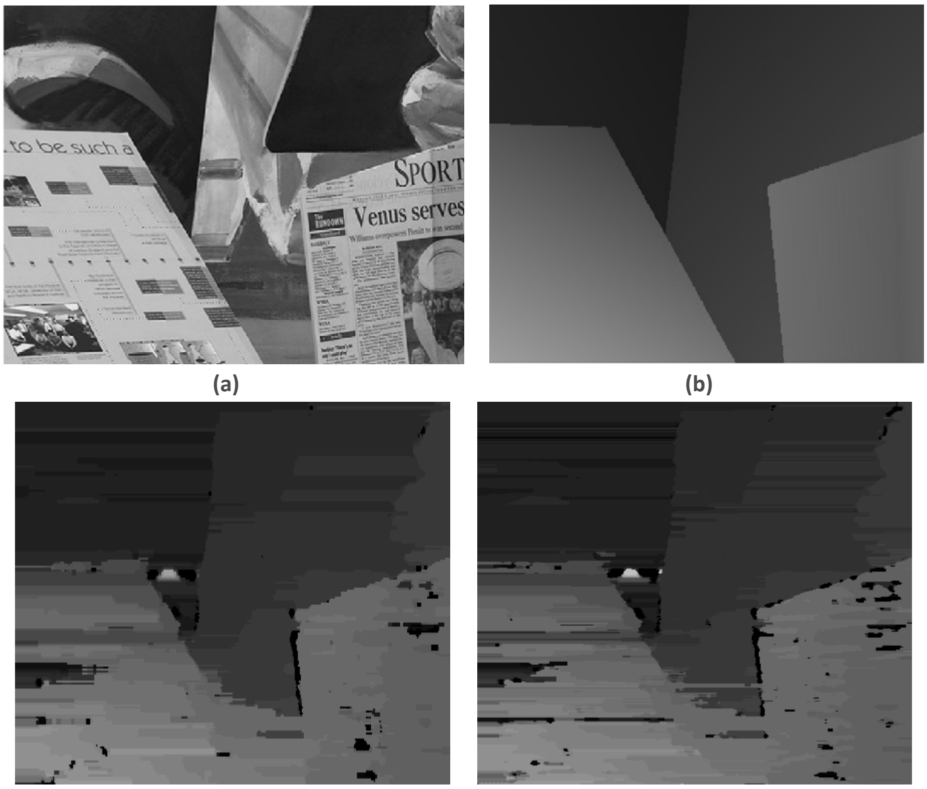
Fig. 5.52. (a) Original Teddy left image, (b) true depth map. Teddy depth map estimation with (a) 187 and (b) 375 scan lines.

The time analysis is shown in table 5.11.

Number of scan lines	Time (ms)	fps	Non-occluded Errors(%)
(a) 187	25.6	39	24.5
(b) 375	52.6	19	25.7

Table 5.11. Performance of the algorithm for different number of scan lines in the Teddy pair of images.

Finally, the Venus pair of images are processed and presented in figure 5.53 and table 5.12. These images have a resolution of 383×434 pixels and a median filter of 5×5 was implemented before presenting.



(a) (b)

(c) (d)

Fig. 5.53. (a) Original Venus left image, (b) true depth map. Venus depth map estimation with (a) 191 and (b) 383 scan lines.

The time analysis is shown in table 5.12.

Number of scan lines	Time (ms)	fps	Non-occluded Errors(%)
(a) 191	17.5	57	11.6
(b) 383	35.7	28	12.4

Table 5.12. Performance of the algorithm for different number of scan lines in the Venus pair of images.

It is important to notice that these two last examples are highly rich in textures, where algorithms dealing with edges or region growing use to have several problems.

Nowadays there is no algorithm that achieves accurate depth maps estimations without investing a huge amount of computational resources and time.

The algorithms of real-time correlation based from [38], (b) a segment-based algorithm from [56] and (c) also segment-based algorithm from [57] will be used for time performance comparisons.

The quality of these depth map estimations is higher than that proposed in this section, but regarding the time performance, we find interesting results. Hirschmüller *et al.* obtain this result over a 450MHz processor at 4.7fps (notice that the image size is 240×320 in this experiment) [38]. Hong and Chen get more accurate with a 2.4GHz PC after 3 seconds (image size 288×384). In the case of Klaus *et al.*, the computation time required is higher than 14 seconds in a 2.21GHz machine [57] (the size is not specified in this work). We can easily see the improvement in terms of performance of our algorithm, since our results for the same image achieve a frame rate of 84fps in slower processors and best case (except the case of Hirschmüller *et al.*).

Regarding to Dynamic Programming Matching algorithms, we can find the following results, shown in figure 5.54.

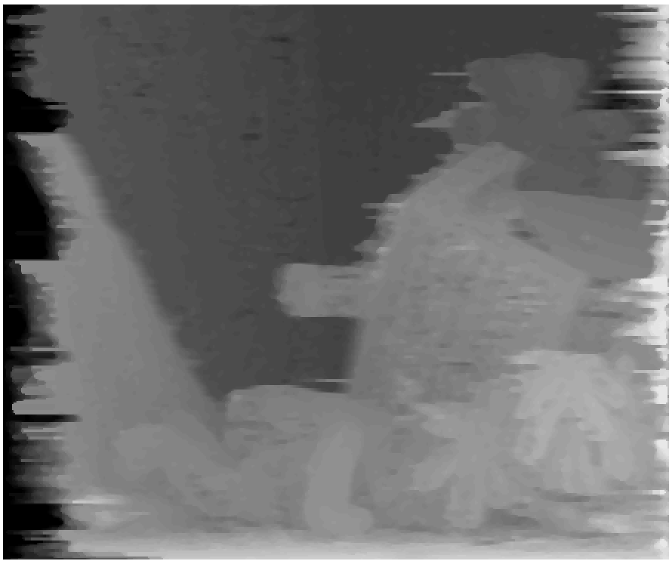


Fig. 5.54. Dynamic programming example of Teddy set, from [79].

The "streaky" lines are evident in figure 5.54, as said before about these algorithms. They are fast (1D analysis) by paying the price of horizontal uncorrelated estimations. In our case, this is attenuated by means of a mean filter over the depth map.

Regarding to the complexity of the algorithm, we can appreciate linearity in the computation time as a function of the number of lines. Moreover, the order of the algorithm we present and analyze in this section is *O(M·L)*, being *M* the number of columns of the image and *L* the number of lines to be scanned. No dependency of the height of the image was found. Just for comparison purposes, we can take the example given in [78]. The algorithm proposed in this section matches points for fingerprint identification. It has a complexity of $O(n^2k^2\log n)$, being *k* the operator size and *n* the size of the image.

This essential difference is due to the highly specific approach adopted in this work: we don't deal with rotation or scale variant images, as is done in [78]. We can assume that the epipolar lines are horizontal and identical in both images and, hence, we don't have to search the feature points in a 2D space. This constraint reduces in one dimension the complexity of the problem. We can exploit all these specification of a specific stereo vision problem to adjust the searching and matching algorithm to make it as simple as possible. Then, its complexity falls down dramatically and although final results are not very accurate (it depends of the final application, as said), we find it interesting as a new way to solve the stereo matching problem.

Finally, we can clearly see the arrangement between complexity and accuracy. In fact, the goodness of each algorithm depends on the final application of the algorithm. What we have searched is a light way to estimate the depth map when accuracy is not critical.

The following video shows a real-life sequence of images, with the depth map extraction around 30 and 50 fps, without median filtering.

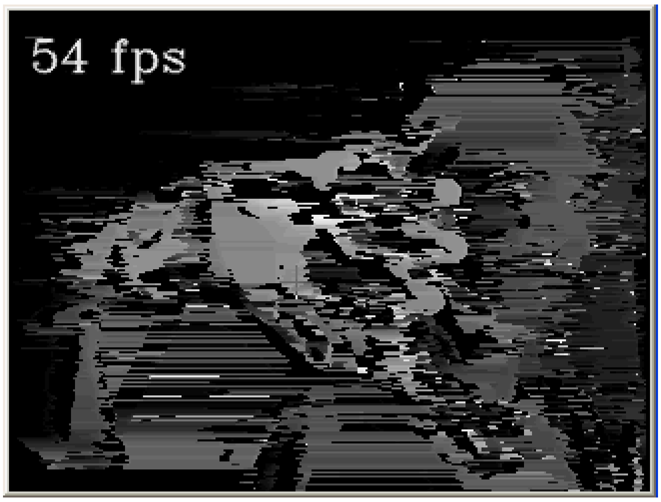
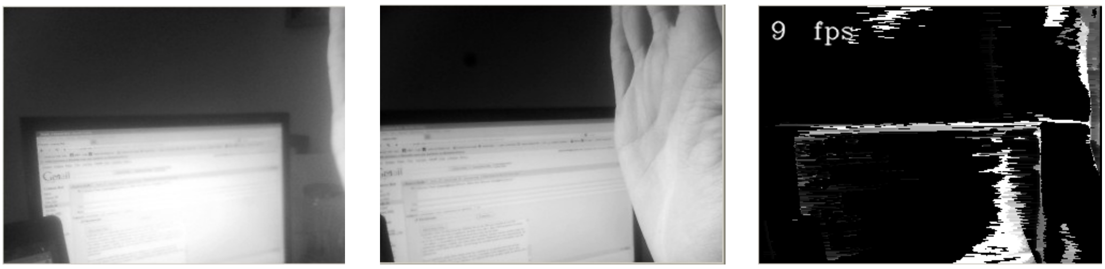


Fig. 5.55. Ctrl+Click on the image to see the video (resource must be available in ./resources).

This last algorithm can be implemented in a memory lite version, processing one row each time, not needing to store the whole images in the RAM. In this version, two rows are read from the image capture devices, and the corresponding 3D row is computed. The accuracy and time performances are the same. Only how the memory is managed changes.

This is an improvement regarding the low-cost hardware implementation, which will implement this algorithm.

However, we found some problems, related with the auto-exposure system of the cameras involved, as shown in figure 5.56.



(a)

(b)

(c)

Fig. 5.56. (a) Left image in a real situation, (b) right image in the same situation and (c) generated depth map.

In these figures, one can see how each camera has implemented a different level of the exposure (see, for example, the background gray level), due to the presence of the hand, much more brilliant and with a bigger area in the right image (thus, the background becomes darker). In fig. 5.56c, we can appreciate the poor estimation of the depth of the screen, for example, because of the explained reason.

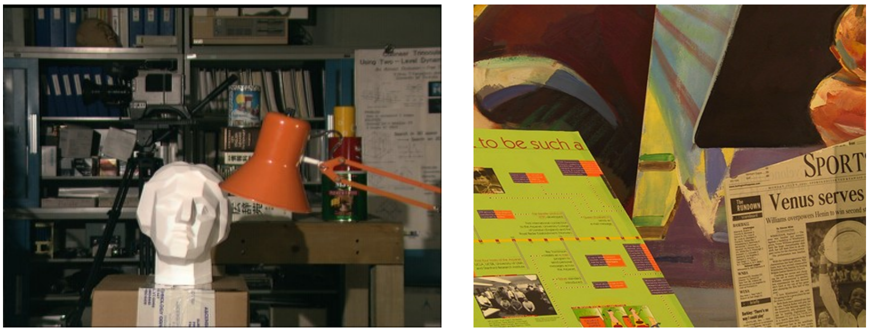
5.5. Algorithm Comparison and Discussion

In this section, we will present a comparison among proposed algorithms.

5.5.1. Accuracy Comparison

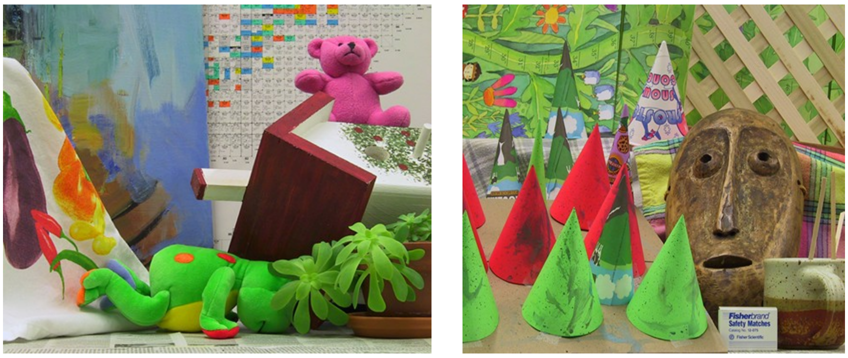
To evaluate proposed algorithms, as it was shown, we used the Middelbury test bed [1]. This is a set of stereo and rectified pair of images, with their ground truth. Moreover, in this webpage there is available an online tool to evaluate new algorithms with 4 different images (in terms of textures, disparities, occlusions, etc.), and to put them in an ordered table with other well known or anonymous algorithms.

The four images used in the benchmark are taken from Middlebury database [1] and presented in figure 5.57 (only left view shown).



(a)

(b)



(c) (d)

Fig. 5.57. (a) Tsukuba, (b) Venus, (c) Teddy and (d) Cones left view, from [1].

To fulfill the table, we will use *full* (F) and *half* (H) definition versions (every line is a scan line and one over two lines is a scanline) in the cases of points and dense matching.

The following table shows the quantitative evaluation of the accuracy for the proposed algorithms:

Images	Tsukuba			Venus			Teddy			Cones			Average
	% of error pixels (threshold of 2%)												
Algorithm	Non-occl	all	disc	Non-occl	all	disc	Non-occl	all	disc	Non-occl	all	disc	
B/W	55.9	**56.6**	66	73.9	**74.2**	68.7	79	**80.7**	87	84.4	**85.8**	93.1	**75.5**
Color	46.9	**47.6**	52.4	77.2	**77.6**	90.2	60.0	**63.2**	69.5	88.7	**89.3**	90.8	**71.1**
Points(F)	11.3	**12.9**	28.7	14.9	**15.7**	42.2	45.4	**49.5**	62.6	64.6	**66.7**	65.7	**40.0**
Points(H)	10.5	**12.1**	31.0	15.2	**16.0**	44.3	45.4	**49.5**	62.6	64.6	**66.7**	65.7	**40.3**
Dense(F)	9.72	**10.6**	27.3	12.4	**13.0**	31.0	25.7	**31.1**	34.7	28.3	**33.4**	40.9	**24.8**
Dense(H)	7.80	**8.74**	26.9	11.6	**12.2**	31.0	24.5	**29.9**	38.6	28.1	**33.1**	41.1	**24.5**

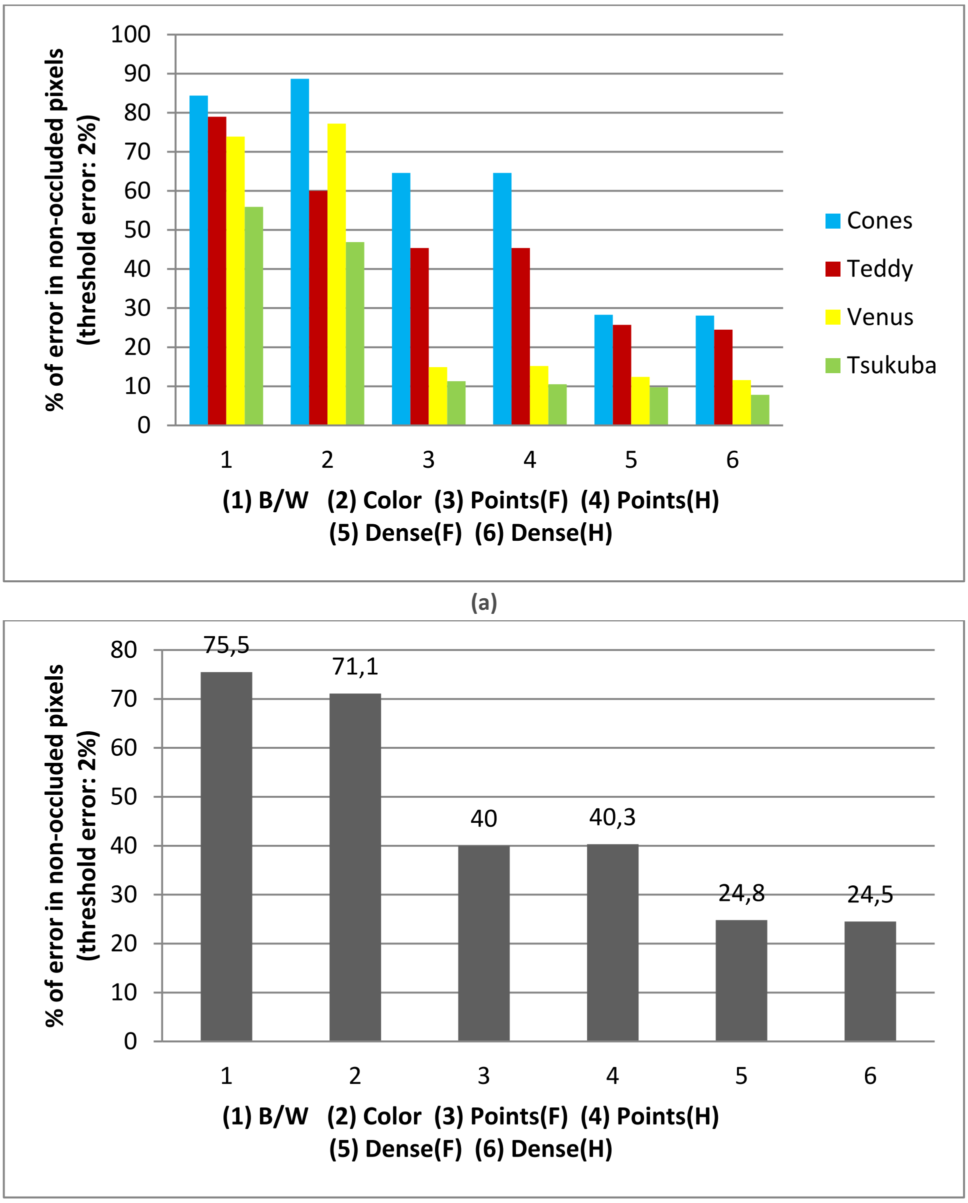
Table 5.13. Accuracy comparison of proposed algorithms.

Segment-based matching has presented some important problem, overall related to textured regions, where the minimum-area condition discards many pixels.

Video applications of segment-based algorithms revealed some extra problems: The matching between characteristics vectors is done via a nested conditionals chain. Hence, with some small differences in segmentation, the same real region can be segmented slightly different in one frame and the following one. Then, the matching process may not find the correspondence. In practical terms, the 3D video extracted with these methods use to blink, matching some regions and not matching them in the following frame and vice versa. This effect was very annoying and made this application useless for our aim.

That was the main reason to try other possibilities.

Next graphs present the evolution of implemented algorithms in terms of accuracy.



(a)

(b)

Fig. 5.58. Evolution, (a) for each image and (b) average errors.

We can easily see the improvement of accuracy. The color algorithm, against the B/W, increases the accuracy of average error pixels, while we can appreciate an increase of errors in images with not too much color information, like Tsukuba and Venus pair of images. In these pairs of images, the B/W algorithm works in a better way. When we arrive to highly color images (Teddy and Cones), the color algorithm shows its advantages.

In any case, the improvement of both points and dense depth map estimation algorithms is evident. Errors are decreased to almost the half in the case of points matching. Regarding the dense algorithm, the decrease achieves a 66% reduction of average total errors.

Analyzing more in detail in terms of images characteristics, we can conclude the following aspects:

- Venus, Teddy and Cones pairs are much richer in textures than Tsukuba pair of images, and region-growing algorithms works in a very deficient way with highly textured images.
- Points, Dense and, in general terms, “differential” algorithms, work much better with this kind of images, as it was explained in the State of the Arte section of this chapter.
- Both region-growing and dense matching algorithms are stable when the texture increases, while points matching algorithm is very sensitive to these changes.
- In Cones and Teddy sets, the difference between region-growing and points matching is not so relevant (however, this difference is much bigger when they are compared with the dense matching algorithm). In the case of Venus and Tsukuba pair of images, this difference is the bigger one among the algorithms.

Finally, we can summarize the obtained data affirming an evident evolution (clearly seen in figure 5.58b) that, for instance, invites us to use the last algorithm presented, the dense and fast depth map algorithm.

5.5.2. Time Consumption Comparison

We have already seen how the proposed algorithms work in terms of accuracy and pixel error. But a very important constraint of our application is time and computational load. I will present in this section a comparison and discussion of the proposed algorithms in terms of time performance. Computing conditions are those of the previous sections.

Table 5.14 presents time consumption for the same cases analyzed before.

Images	Tsukuba 288×384	Venus 434×383	Teddy 375×450	Cones 375×450	Average time/pixel
Algorithm	Time consumption [ms] (frames per second)				[ns]
B/W	50 (20)	76.6 (13)	78.9 (12)	78.6 (13)	462.5
Color	77.4 (12)	111.8 (8)	114.2 (8)	114.5 (8)	680.3
Points(F)	23.2 (43)	35.7 (28)	41.6 (24)	40.9 (24)	230.2
Points(H)	11.9 (84)	17.5 (57)	20 (50)	24.4 (41)	120.1
Dense(F)	13.8 (72)	26.3 (38)	43.5 (23)	76.5 (13)	260.6
Dense(H)	6.9 (145)	14.9 (67)	25.6 (39)	40 (25)	142.3

Table 5.14. Time comparison of proposed algorithms.

Since in the set of images, there are different sizes, it is presented in the previous table the average of the time used per pixel of original image (just one of them). We see in this comparison the evolution of efficiency achieved.

Regarding to the B/W algorithm with the color one, we see an increase of computation time. This effect is justified because color images, as said before, have three times more information. However, we can appreciate that the increase of time is not by a factor of 3, but around 50%. This was the main advantage of this new algorithm: increasing the accuracy without increasing a lot the computation time.

We can see, as done in the previous section, the big difference with the two other proposals, points and dense matching.

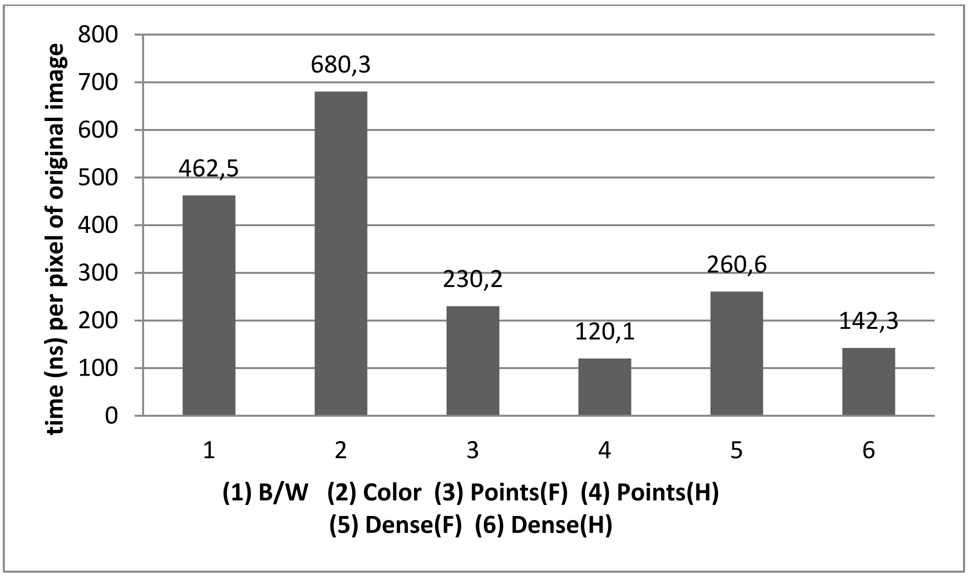


Fig. 5.59. Evolution of computation time per pixel.

We can appreciate a slight increase of computation time from points matching to dense matching. Analyzing the specific time required to process each image, we see that the responsible of this soft increase is the Cones image. While in the rest of images the dense matching algorithm achieves the same time, or even smaller, for this last image there is an increase of 50% in both cases, full and half resolution. Both points and dense matching algorithms are *content dependent,* the computation time depends of the internal organization of the image, and not only of the image size. Moreover, this dependency is different between them, as we can appreciate in the previous table.

5.5.3. Memory Consumption Comparison

The final parameter to be analyzed is the memory used by the different algorithms. Memory is an important constraint, since cheap microprocessors or microcontrollers use to have several limitations in memory offered for program data, stored in the RAM.

Table 5.15 shows the size of the executable of each algorithm, with the size of the used data during execution. This last parameter is given as a proportional value to the original gray scale image size (N×M). I offer also an example of total user memory required with the bigger image processed in this work.

Algorithm	Executable size [bytes]	Data size [bytes]	Data size example (375×450) [MB]
B/W	68K	12·N·M+223	2.25
Color	72K	16·N·M+223	2.93
Points(F)	52K	3·N·M+2100·M+28	1.29
Points(H)	52K	3·N·M/2+2100·M/2+28	0.65
Dense(F)	52K	3·N·M+16	0.51
Dense(H)	52K	3·N·M/2+16	0.25
DenseLite(F)	52K	N·M+2·M+16	0.17
DenseLite(H)	52K	N·M/2+2·M+16	0.09

Table 5.15. Memory comparison of proposed algorithms.

It is clear, regarding to the data offered in this table, that the simplicity of each algorithm code (the size of the executable file) does not change a lot from one implementation to another. The most important difference is achieved comparing the color algorithm with the points or the dense matching algorithm. This difference is around 50% more in the case of the color one.

Huge differences appear when comparing the consumed memory. It is graphically shown in the following graph.

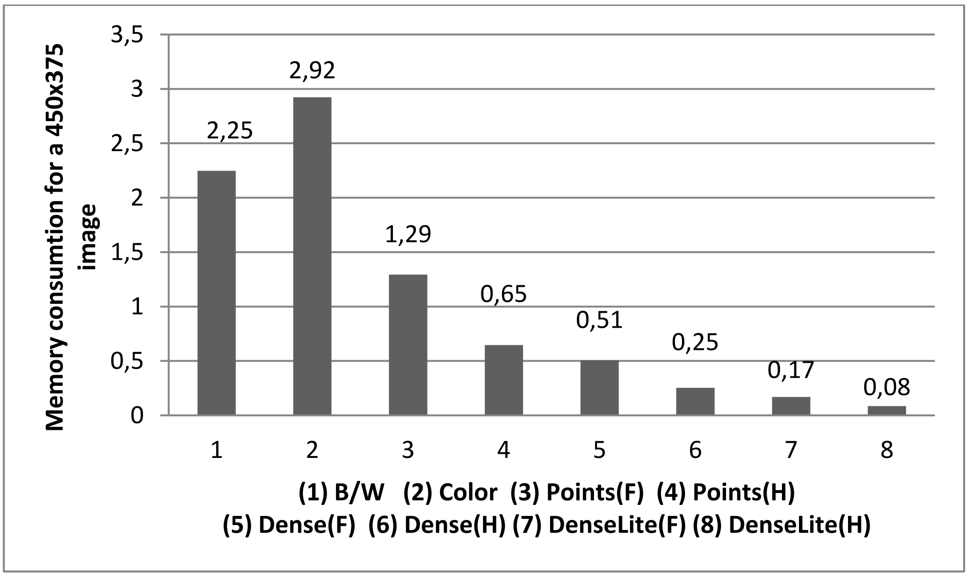


Fig. 5.60. Memory consumptions of proposed algorithms.

5.5.4. Algorithms Comparison

Even though we can guess that there are some algorithms from those presented in this chapter that are better (in several senses) than others, I will perform a final comparison among them, by means of a unified index:

$$i = \frac{1}{error \cdot time \cdot memory} \tag{5.16}$$

With this index, which punishes error, time and memory in the same way, we can compare in one single graph the global performance of each algorithm. The higher this index is, the better the algorithm works.

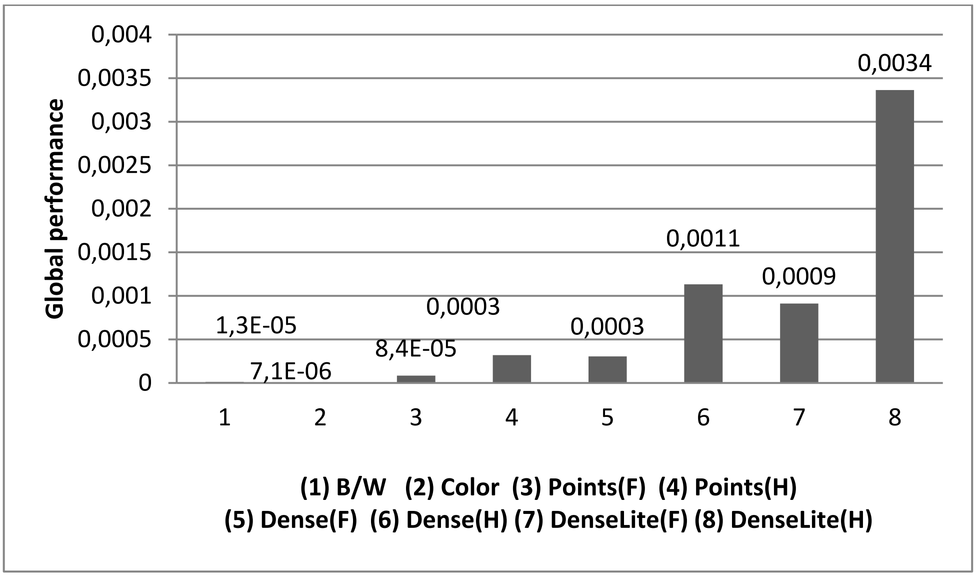


Fig. 5.61. Global comparison of proposed algorithms.

The best algorithm has a global better performance in 4 magnitude orders regarding the worst one. This algorithm will be, then, used in further parts of this work.

5.5.5. Comparison with the State of the Art

Once we can choose an implementation from those presented in this chapter, I will perform a comparison with some other relevant algorithms published.

The main problem that we find when trying to perform such a task is the deficient information provided sometimes in literature. This lack is due to the specific objectives of each publication. Some of them, just focus on accuracy, hence, time is not important and, even, avoidable.

I chose, then, some implementations with which compare this algorithm in the same terms (i.e., measured with the same tool). The selection is done over the best performance algorithm in the Middlebury classification [80].

Another important drawback found when comparing algorithms is the differences among the machines used to implement each algorithm. In some cases (as in [81]), we see specific hardware like GPU's implementing this algorithm to achieve real-time performances. However, there is no other way, for instance, to compare these algorithms. We just propose a qualitative comparison based on computing conditions and retrieved computation time.

The time is normalized per pixel, because different images (Teddy and Tsukuba, for example) have different sizes. This normalized time is presented, as it was done when comparing the proposal presented in this work, in nanoseconds/pixel.

The error is the average error for all images with a threshold of 2. In the comment column is presented some relevant information about the computation conditions, as well as the bibliography source.

Table 5.16 presents some interesting results found in newest publications in this field.

Algorithm	Error	Time [s]	Time[ns]/pixel	Comments
1 [82]	5.18	32	189629.6	Time for Teddy images CPU @2.14GHz
2 [81]	6.46	6	54253.5	Time for Tsukuba images GPU @3GHz
3 [83]	2.85	20	118518.5	Time for Tsukuba images CPU @1.6GHz
4 [84]	5.88	.6	3555.5	Time for Teddy images CPU @2.14GHz Duo
5 [85]	4.86	.465	4205.7	Time for Tsukuba images CPU @2.83GHz
6 [86]	2.72	20	180844.9	Time for Tsukuba images CPU @N/A
7 [87]	2.54	14	126591.4	Time for Tsukuba images CPU @2.21GHz
8 [88]	3.01	600	5425347.2	Time for Tsukuba images CPU @N/A
Dense (F)	10.1	0.007	260	Time for Tsukuba images CPU @1.6GHz

Table 5.16. Different algorithms with the corresponding results.

There are some important points to be remarked in this table. In the case of the second algorithm, it runs over a GPU, specifically designed hardware to manage images, with a massive parallel architecture which works much faster than general purpose CPU's.

We can appreciate this difference in the following figure, extracted from [79]:

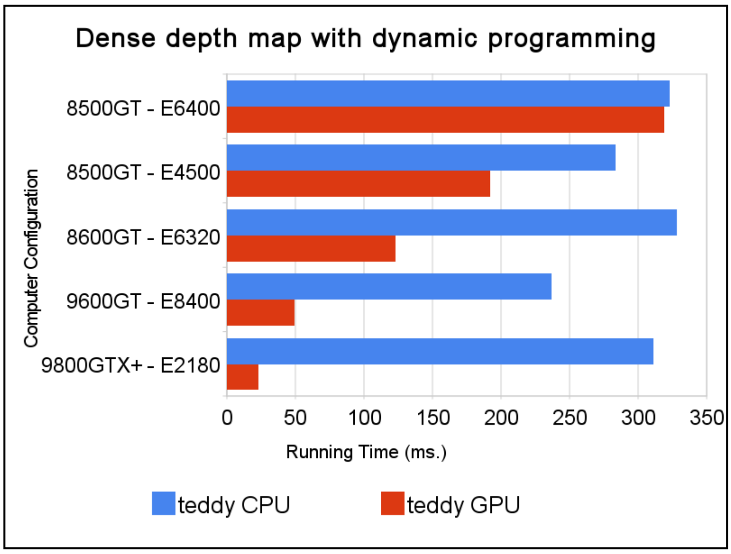


Fig. 5.62. Teddy computation times for GPU and CPU.

There are some outlier results, like algorithm 8, which takes around 10 minutes to obtain some of the most accurate results. Finally, let's point out that some papers provide computation time, but not the computing conditions.

As done before when comparing a global descriptor of these algorithms, we can compute an index (just taking into account time and errors, because the memory is not given in any case).

Algorithm	Index
1 [82]	1.01804E-06
2 [81]	2.85325E-06
3 [83]	2.96053E-06
4 [84]	4.78316E-05
5 [85]	4.89244E-05
6 [86]	2.03294E-06
7 [87]	3.11001E-06
8 [88]	6.12359E-08
9 [OUR PROPOSAL]	2.92809E-04

Table 5.17. Global index of compared algorithms.

Even if that comparison cannot be done directly as proposed, because of the different machines, we can take it as a "worst case" comparison, since the computer used to measure our proposal is the slowest one of those specified in the articles.

Figure 5.63 represents a graphical comparison of the 9 compared algorithms in terms of this global merit index.

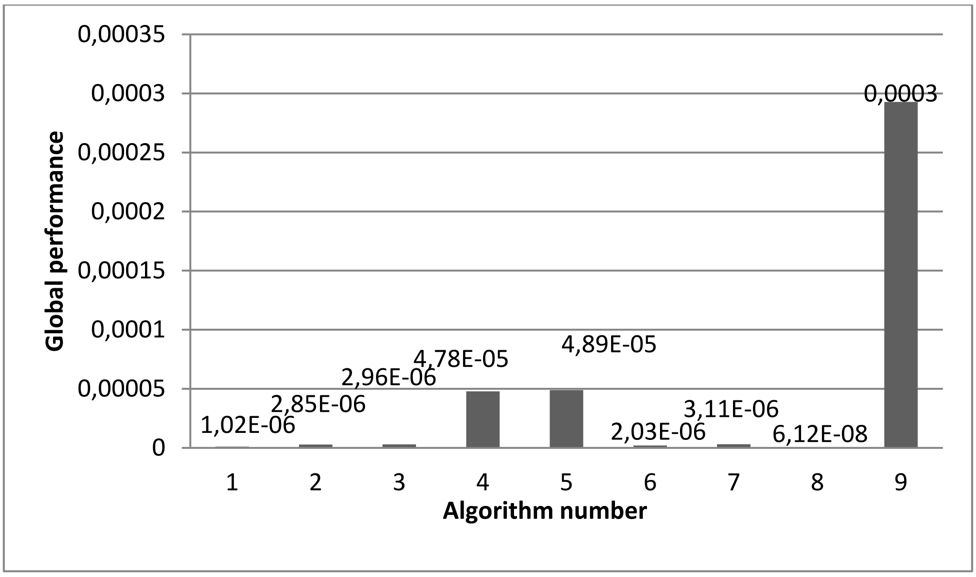


Fig. 5.63. Comparison among different algorithms and the proposed one.

As it can be seen, one magnitude order comes between the best algorithm (presented in [85]), measured with the proposed index, and ours. This is done not by the accuracy (where our algorithm is the worst one) but in the global performance, since we are interested in a very fast implementation with the counterpart of a decrease of accuracy.

5.6. Conclusions

In this chapter we have reviewed some of the most important techniques and algorithms in the field of the image processing. Through this revision, we have been able to appreciate the depth and complexity of this research area, which is extremely active nowadays.

Likewise, we have proposed 4 different approaches to solve the disparity and depth extraction problem, with increasing performances, measured with a merit index described in eq. 5.16. The last version of these algorithms is likely prepared to work in low-cost and low-power hardware and, hence, useful for the project.

In chapter 1, H2 hypothesis stated that light and fast image processing algorithms can be designed, and we can affirm that this assumption has been proved.

In the proposal chapter 4, general and specific goals were described and stated. Concerning those about the image processing, we can extract the following:

- Memory less than 307KB
- Performance over 24fps
- Accuracy of 75%

Several algorithms were proposed and evaluated. In the final prototype, in both the software and hardware versions, we used the last one, given that its performances were much higher than the others. However, we had several options in this case, since the algorithm allowed us processing all the lines, or just a fraction of them.

Regarding the memory use, it is shown that the full version of the algorithm used 230.4KB (corrected for a reference image of 320×240, instead of 375×450).

The time performance was measured over static images on a CPU 1.6GHz, obtaining 19ms in average and, hence, 54fps.

Another condition established in the proposal chapter was the accuracy of the image processing algorithm. Although it is variable regarding the image itself, an average accuracy over static images (the only one we could measure, since we have no depth truth references for our real life images) was calculated in 75.2%.

These results and their comparison with well-established proposals have yield to the following peer reviewed publications in journals, conference proceedings and international books:

- Authors: Pablo Revuelta Sanz, Belén Ruiz Mezcua, José M. Sánchez Pena
  Title: Estimating Complexity of Algorithms as a Black-Box Problem: A Normalized Time Index.
  Conference: 3rd International Multi-Conference on Complexity, Informatics and Cybernetics (IMCIC 2012).
  Publication: (not available yet).
  Place: Orlando, Florida (U.S.A.).
  Date: March 25th-28th, 2012.

- Authors: Pablo Revuelta Sanz, Belén Ruiz Mezcua, José M. Sánchez Pena
  Title: Stereo Vision Matching over Single-Channel Color-Based Segmentation

Conference: International Conference on Signal Processing and Multimedia Applications (SIGMAP 2011)
Publication: Proceedings SIGMAP 2011, pp. 126-130.
Place: Seville (Spain).
Date: July, 2011.

- Authors: Pablo Revuelta Sanz, Belén Ruiz Mezcua, José M. Sánchez Pena
  Title: Efficient Characteristics Vector Extraction Algorithm using Auto-seeded Region-Growing.
  Conference: 9th IEEE/ACIS International Conference on Computer and Information Science (ICIS 2010)
  Publication: Proceedings of the 9th IEEE/ACIS International Conference on Computer and Information Science (ICIS 2010), pp. 215-221.
  Place: Kaminoyama (Japan).
  Date: August, 2010.

- Authors: Pablo Revuelta, Belén Ruiz Mezcua, José Manuel Sánchez Pena, Jean-Phillippe Thiran.
  Title: Segment-Based Real-Time Stereo Vision Matching using Characteristics Vectors
  Journal: Journal of Imaging Science and Technology 55(5)
  Date: Sept./Oct. 2011.
  Selected feature article of the volume.

- Authors: Revuelta Sanz, P., Ruiz Mezcua, B., & Sánchez Pena, J. M.
  Title: Depth Estimation. An Introduction.
  Book title: Current Advancements in Stereo Vision. 224 pp.
  Editor: Ms. Marina Kirincic. InTech Ed. In press.
  Place: Rijeka, Croatia.
  Date: 2012
  ISBN: 978-953-51-0660-9, ISBN 979-953-307-832-7
  http://www.intechopen.com/books/export/citation/EndNote/current-advancements-in-stereo-vision/depth-estimation-an-introduction

Finally, there is a submitted paper, currently under review:

- Authors: Pablo Revuelta, Belén Ruiz Mezcua, José Manuel Sánchez Pena
  Title: Fast and Dense Depth Map Estimation for Stereovision Low-cost Systems
  Journal: Image Science Journal
  Date: Feb. 2013.

References

1. D. Scharstein, "Middlebury Database." . 2010.

2. M. Bleyer and M. Gelautz, "A layered stereo matching algorithm using image segmentation and global visibility constraints." *Journal of Photogrammetry & Remote Sensing* vol. 59, pp. 128-150. 2005.

3. J. Kostková and R. Sára, "Fast Disparity Components Tracing Algorithm for Stratified Dense Matching Approach." vol. Research Reports of CMP, Czech Technical University in Prague, No. 28, 2005. 2006.

4. Ahsmail, "Human Vision System." . 2010.

5. M. Bleyer, "Segmentation-based Stereo and Motion with Occlusions,", Institute for Software Technology and Interactive Systems, Vienna University of Technology, 2006.

6. D. L. Pham, C. Xu, and J. L. Prince, "Current Methods in Medical Image Segmentation." *Annual Review of Biomedical Engineering* vol. 2, pp. 315-337. 2000. Annual Reviews Inc.

7. S. Zhao, "Image Registration by Simulating Human Vision." *Lecture Notes in Computer Science, Advances in Image and Video Technology* vol. 4872, pp. 692-701. 2007. Berlin Heidelberg, Springer-Verlag.

8. J. Turski, "Computational harmonic analysis for human and robotic vision systems." *Neurocomputing* vol. 69, pp. 1277-1280. 2006. Elsevier B.V.

9. N. Ouerhani, A. Bur, and H. Hügli, "Linear vs. Nonlinear Feature Combination for Saliency Computation: A Comparison with Human Vision." *Lecture Notes in Computer Science, Pattern Recognition* vol. 4174, pp. 314-323. 2006. Berlin Heidelberg, Springer-Verlag.

10. I. Kurki and J. Saarinen, "Shape perception in human vision: specialized detectors for concentric spatial structures?" *Neuroscience Letters* vol. 360, pp. 100-102. 2004.

11. T. S. Meese and R. J. Summers, "Area summation in human vision at and above detection threshold." *Proceedings of the Royal Society B: Biological Sciences* vol. 274, pp. 2891-2900. 2009.

12. G. H. Jacobs, G. A. Williams, H. Cahill et al., "Emergence of Novel Color Vision in Mice Engineered to Express a Human Cone Photopigment." *Science* vol. 315, pp. 1723-1727. 2007.

13. C. F. Stromeyer, R. E. Kronauer, J. C. Madsen et al., "Opponent-Movement Mechanisms in Human-Vision." *Journal of the Optical Society of America A-Optics Image Science and Vision* vol. 1 no. 8, pp. 876-884. 1984.

14. K. Racheva and A. Vassilev, "Human S-Cone Vision Effect of Stiumuls Duration in the Increment and Decrement Thresholds." *Comptes rendus de l'Academie bulgare des Sciences* vol. 62 no. 1, pp. 63-68. 2009.

15. S. Guttman, L. A. Gilroy, and R. Blake, "Spatial grouping in human vision: Temporal structure trumps temporal synchrony." *Vision Research* vol. 47, pp. 219-230. 2007.

16. E. Levinson and R. Sekuler, "Independence of Channels in Human Vision Selective for Direction of Movement." *Journal of Physiology-London* vol. 250 no. 2, pp. 347-366. 1975.

17. M. Georgeson, "Antagonism between Channels for Pattern and Movement in Human Vision." *Nature* vol. 259 no. 5542, pp. 412-415. 1976.

18. A. Saxena, S. H. Chung, and A. Y. Ng, "3-D Depth Reconstruction from a Single Still Image." *International Journal of Computer Vision* vol. 76, pp. 53-69. 2008.

19. T. S. Douglas, S. E. Solomonidis, W. A. Sandham et al., "Ultrasound image matching using genetic algorithms." *Medical and Biological Engineering and Computing* vol. 40, pp. 168-172. 2002.

20. L. Yao, L. Ma, and D. Wu, "Low Cost 3D Shape Acquisition System Using Strip Shifting Pattern." *Lecture Notes in Computer Science, Digital Human Modeling* vol. 4561, pp. 276-285. 2007. Berlin Heidelberg, Springer-Verlag.

21. J. P. O. Evans, "Stereoscopic imaging using folded linear dual-energy x-ray detectors." *Measurement Science and Technology* vol. 13, pp. 1388-1397. 2002. U. K., INSTITUTE OF PHYSICS Publishing Ltd.

22. A. R. J. François and G. G. Medioni, "Interactive 3D model extraction from a single image." *Image and Vision Computing* vol. 19, pp. 317-328. 2001. Elsevier Science B.V.

23. K. E. Ozden, K. Schindler, and L. van Gool, "Simultaneous Segmentation and 3D Reconstruction of Monocular Image Sequences." *Computer Vision, 2007.ICCV 2007.IEEE 11th International Conference on* , pp. 1-8. 2007.

24. F. S. Helmi and S. Scherer, "Adaptive Shape from Focus with an Error Estimation in Light Microscopy." *2nd Int'l Symposium on Image and Signal Processing and Analysis* , pp. 188-193. 2001.

25. A. S. Malik and T.-S. Choi, "Depth Estimation by Finding Best Focused Points Using Line Fitting." *Lecture Notes in Computer Science, Image and Signal Processing* vol. 5099, pp. 120-127. 2008.

26. Y. Jia, Y. Xu, W. Liu et al., "A Miniature Stereo Vision Machine for Real-Time Dense Depth Mapping." *Lecture Notes in Computer Science, Computer Vision Systems* vol. 2626, pp. 268-277. 2003. Berlin Heidelberg, Springer-Verlag.

27. J.-P. Pons and R. Keriven, "Multi-View Stereo Reconstruction and Scene Flow Estimation with a Global Image-Based Matching Score." *International Journal of Computer Vision* vol. 72 no. 2, pp. 179-193. 2007.

28. H. K. I. Kim, K. Kogure, and K. Sohn, "A Real-Time 3D Modeling System Using Multiple Stereo Cameras for Free-Viewpoint Video Generation." *Lecture Notes in Computer Science, Image Analysis Recognition* vol. 4142, pp. 237-249. 2006. Berlin Heidelberg, Springer-Verlag.

29. S. M. Seitz and J. Kim, "The Space of All Stereo Images." *International Journal of Computer Vision* vol. 48 no. 1, pp. 21-38. 2002. The Netherlands, Kluwer Academic.

30. T. Tuytelaars and L. v. Gool, "Matching Widely Separated Views Based on Affine Invariant Regions." *International Journal of Computer Vision* vol. 59 no. 1, pp. 61-85. 2004.

31. V. N. Radhika, B. Kartikeyan, G. Krishna et al., "Robust Stereo Image Matching for Spaceborne Imagery." *IEEE Transactions on Geoscience and Remote Sensing* vol. 45 no. 9, pp. 2993-3000. 2007.

32. D. Scharstein and R. Szeliski, "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms." *International Journal of Computer Vision* vol. 47 no. 1, pp. 7-42. 2002.

33. C. Schimd, A. Zisserman, and R. Mohr, "Integrating Geometric and Photometric Information for Image Retrieval." *Lecture Notes in Computer Science, Shape, Contour and Grouping in Computer Vision* vol. 1681, pp. 217-233. 1999.

34. Ch. Wang and M. L. Gavrilova, "A Novel Topology-Based Matching Algorithm for Fingerprint Recognition in the Presence of Elastic Distortions." *Lecture Notes in Computer Science, Computational Science and Its Applications ICCSA* vol. 3480, pp. 748-757. 2005.

35. Z. He and Q. Wang, "A Fast and Effective Dichotomy Based Hash Algorithm for Image Matching." *Lecture Notes in Computer Science, Advances in Visual Computing* vol. 5358, pp. 328-337. 2009.

36. M. S. Islam and L. Kitchen, "Nonlinear Similarity Based Image Matching." *International Federation for Information Processing* vol. 228, pp. 401-410. 2004.

37. J. Williams and M. Bennamoun, "A Non-linear Filtering Approach to Image Matching." *Proceedings of the 14th International Conference on Pattern Recognition* vol. 1 no. 1, p.3. 1998.

38. H. Hirchsmüller, P. R. Innocent, and J. Garibaldi, "Real-Time Correlation-Based Stereo Vision with Reduced Border Errors." *Journal of Computer Vision* vol. 47 no. 1/2/3, pp. 229-246. 2002.

39. L. Tang, Ch. Wu, and Z. Chen, "Image dense matching based on region growth with adaptive window." *Pattern recognition letters* vol. 23, pp. 1169-1178. 2002.

40. J. Yu, L. Weng, Y. Tian et al., "A Novel Image Matching Method in Camera-calibrated System." *Cybernetics and Intelligent Systems, 2008 IEEE Conference on* , pp. 48-51. 2008.

41. K. Li, S. Wang, M. Yuan et al., "Scale Invariant Control Points Based Stereo Matching for Dynamic Programming." *The Ninth International Conference on Electronic Measurement & Instruments (ICEMI'2009)* , pp. 3-769-3-774. 2009.

42. H. Lang, Y. Wang, X. Qi et al., "Enhanced point descriptors for dense stereo matching." *Image Analysis and Signal Processing (IASP), 2010 International Conference on* , pp. 228-231. 2010.

43. B. Liu, H.-B. Gao, and Q. Zhang, "Research of Correspondence Points Matching on Binocular Stereo Vision Measurement System Based on Wavelet." *Machine Learning and Cybernetics, 2006 International Conference on* , pp. 3687-3691. 2006.

44. J. C. Kim, K. M. Lee, B. T. Choi et al., "A dense stereo matching using two-pass dynamic programming with generalized ground control points." *Computer Vision and Pattern*

Recognition, 2005.CVPR 2005.IEEE Computer Society Conf.on vol. 2, pp. 1075-1082. 2005.

45. H. Marr and T. Poggio, "Cooperative computation of stereo disparity." *Science* vol. 194, pp. 283-287. 1976.

46. H. Mayer, "Analysis of means to improve cooperative disparity estimation." *ISPRS Conference on Photogrammetric Image Analysis* vol. XXXIV. 2003.

47. L. Zitnick and T. Kanade, "A cooperative algorithm for stereo matching and occlusion detection." *IEEE Transactions on Pattern Analysis and Machine Inteligence* vol. 22 no. 7, pp. 675-684. 2000.

48. J. Käck, "Robust Stereo Correspondence using Graph Cuts,", Royal Institute of Technology, 2004.

49. Y. Ohta and T. Kanade, "Stereo by intra- and inter-scanline search using dynamic programming." *IEEE Transactions on Pattern Analysis and Machine Inteligence* vol. 7 no. 2, pp. 139-154. 1985.

50. A. Bobick and S. Intille, "Large occlusion stereo." *International Journal of Computer Vision* vol. 33 no. 3, pp. 181-200. 1999.

51. V. Kolmogorov and R. Zabih, "Computing visual correspondence with occlusions via graph cuts." vol. Technical Report CUCS-TR-2001-1838, Cornell Computer Science Department. 2010.

52. L. Szumilas, H. Wildenauer, and A. Hanbury, "Invariant Shape Matching for Detection of Semi-local Image Structures." *Lecture Notes in Computer Science, Image Analysis Recognition* vol. 5627, pp. 551-562. 2009.

53. Y. Xia, A. Tung, and Y. W. Ji, "A Novel Wavelet Stereo Matching Method to Improve DEM Accuracy Generated from SPOT Stereo Image Pairs." *International Geoscience and Remote Sensing Symposium* vol. 7, pp. 3277-3279. 2001.

54. G. Pajares, J. M. Cruz, and J. A. López-Orozco, "Relaxation labeling in stereo image matching." *Pattern recognition* vol. 33, pp. 53-68. 2000.

55. J. Sun, H.-Y. Shum, and N.-N. Zheng, "Stereo matching using belief propagation." *European Conference on Computer Vision* , pp. 510-524. 2002.

56. L. Hong and G. Chen, "Segment-based Stereo Matching Using Graph Cuts." *Computer Vision and Pattern Recognition, 2004.CVPR 2004.Proc.of the 2004 IEEE Computer Society Conf.on* vol. 1, p.I-74-I-81. 2004.

57. A. Klaus, M. Sormann, and K. Kraner, "Segment-Based Stereo Matching Using Belief Propagation and a Self-Adapting Dissimilarity Measure." *Pattern Recognition, 2006.ICPR 2006.18th International Conference on* , pp. 15-18. 2006.

58. Q. Yu and D. A. Clausi, "SAR Sea-Ice Image Analysis Based on Iterative Region Growing Using Semantics." *IEEE Transaction on Geoscience and Remote Sensing* vol. 45 no. 12, pp. 3919-3931. 2007. IEEE.

59. J. Dehmeshki, H. Amin, M. Valdivieso et al., "Segmentation of Pulmonary Nodules in Thoracic CT Scans: A Region Growing Approach." *IEEE Transactions on Medical Imaging* vol. 27 no. 4, pp. 467-480. 2008. Utrecht, The Netherlands, IEEE.

60. K. Zhang, H. Xiong, X. Zhou et al., "A 3D Self-Adjust Region Growing Method for Axon Extraction." *Image Processing, 2007.ICIP 2007.IEEE International Conference on* vol. 2 no. II, pp. 433-436. 2007. San Diego, California, IEEE Signal Processing Society.

61. L. Gao, J. Jiang, and S. Y. Yang, "Constrained Region-Growing and Edge Enhancement Towards Automated Semantic Video Object Segmentation." *Lecture Notes in Computer Science, Advanced Concepts for Intelligent Vision Systems* vol. 4179, pp. 323-331. 2006. Berlin Heidelberg, J. Blanc-Talon et al. (Eds.).

62. J. Magarey and A. Dick, "Multiresolution Stereo Image Matching Using Complex Wavelets." *Pattern Recognition, 1998. Proceedings of the Fourteenth International Conference on* vol. 1, pp. 4-7. 1998.

63. G. M. Espindola, G. Camara, I. A. Reis et al., "Parameter selection for region-growing image segmentation algorithms using spatial autocorrelation." *International Journal of Remote Sensing* vol. 27 no. 14, pp. 3035-3040. 2006. U. K., Taylor & Francis.

64. Intel Corporation, "Motion Analysis and Object Tracking." *Computer Vision Library*. no. 5, pp. 2-11-2-15. 2001.

65. U. Roeber, E. M. Y. Wong, and A. W. Freeman, "Cross-orientation interactions in human vision." *Journal of Vision* vol. 8 no. 3, pp. 1-11. 2008.

66. P. Revuelta Sanz, B. Ruiz Mezcua, and J. M. Sánchez Pena, "Efficient Characteristics Vector Extraction Algorithm using Auto-seeded Region-Growing." *Proceedings of the 9th IEEE/ACIS International Conference on Computer and Information Science ICIS 2010* , pp. 215-221. 2010.

67. G. Egnal and R. P. Wildes, "Detecting Binocular Half-Occlusions: Empirical Comparisons of Five Approaches." *IEEE Transactions on Pattern Analysis and Machine Intelligence* vol. 24 no. 8, pp. 1127-1133. 2002. IEEE.

68. S.-K. Han, M.-H. Jeong, S.-H. Woo et al., "Architecture and Implementation of Real-Time Stereo Vision with Bilateral Background Subtraction." *Lecture Notes in Computer Science, Advanced Intelligent Computing Theories and Applications.With Aspects of Theoretical and Methodological Issues* vol. 4681, pp. 906-912. 2007. Berlin Heidelberg, Springer-Verlag.

69. Y.-L. Chang and X. Li, "Adaptive Image Region-Growing." *IEEE Transactions on Image Processing* vol. 3 no. 6, pp. 868-872. 1994.

70. M. S. Millán and E. Valencia, "Color image sharpening inspired by human vision models." *Applied Optics* vol. 45 no. 29, pp. 7684-7697. 2006.

71. F. A. A. Kingdom, "Color brings relief to human vision." *Nature Neuroscience* vol. 6 no. 6, pp. 641-644. 2003.

72. J. Nathans, "The Evolution and Physiology of Human Color Vision: Insights from Molecular Genetic Studies of Visual Pigments." *Neuron* vol. 24, pp. 299-312. 1999.

73. R. González and R. E. Woods, *Digital Image Processing,* 2, p. 295: Prentice Hall Press, 2002.

74. Y. Zhang and Y. J. Gerbrands, "Method for matching general stereo planar curves." *Image and Vision Computing* vol. 13 no. 8, pp. 645-655. 1995.

75. J. A. Denton and J. R. Beveridge, "An algorithm for projective point matching in the presence of spurious points." *Pattern recognition* vol. 40 no. 2, pp. 586-595. 2007.

76. R. Elias, "Sparse view stereo matching." *Pattern recognition letters* vol. 28 no. 13, pp. 1667-1678. 2007.

77. W. Lian, L. Zhang, Y. Liang et al., "A quadratic programming based cluster correspondence projection algorithm for fast point matching." *Computer Vision and Image Understanding* vol. 114 no. 3, pp. 322-333. 2010.

78. A. Bishnu, S. Das, S. C. Nandy et al., "Simple algorithms for partial point set pattern matching under rigid motion." *Pattern recognition* vol. 39 no. 9, pp. 1662-1671. 2006.

79. J. Congote, J. Barandiaran, and O. Ruiz, "Realtime Dense Stereo Matching with Dynamic Programming in CUDA." *CEIG'09* . 2009.

80. Middlebury, "Middlebury Evaluation Webpage." . 2012.

81. L. Wang, M. Liao, M. Gong et al., "High-quality real-time stereo using adaptive cost aggregation and dynamic programming." *Third International Symposium on 3D Data Processing, Visualization, and Transmission 2006* , pp. 1-8. 2006.

82. S. MatToccia, S. Giardino, and A. Gambini, "Accurate and efficient cost aggregation strategy for stereo correspondence based on approximated joint bilateral filtering." *Lecture Notes in Computer Science, 2010* vol. 5995/2010, pp. 371-380. 2009.

83. Z. Wang and Z. Zheng, "A region based stereo matching algorithm using cooperative optimization." *Computer Vision and Pattern Recognition, 2008.CVPR 2008.IEEE Conference on* , pp. 1-8. 2010.

84. F. Tombari, S. Mattocia, L. Di Stefano et al., "Near real-time stereo based on effective cost aggregation." *Pattern Recognition, 2008.ICPR 2008.19th International Conference on* , pp. 1-4. 2008.

85. S. Kosov, T. Thormählen, and H.-P. Seidel, "Accurate real-time disparity estimation with variational methods." *ISVC '09 Proceedings of the 5th International Symposium on Advances in Visual Computing: Part I* vol. 5875, pp. 796-807. 2009.

86. Q. Yang, L. Wang, R. Yang et al., "Stereo matching with color-weighted correlation, hierarchical belief propagation and occlusion handling." *IEEE Transactions on Pattern Analysis and Machine Inteligence* vol. 31 no. 3, pp. 1-13. 2009.

87. A. Klaus, M. Sormann, and K. Karner, "Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure." *ICPR '06 Proceedings of the 18th International Conference on Pattern Recognition* vol. 3, pp. 15-18. 2006.

88. M. Bleyer, M. Gelautz, C. Rother et al., "A stereo approach that handles the matting problem via image warping." *2009 IEEE Conference on Computer Vision and Pattern Recognition* , pp. 501-508. 2009.

6. Sonification

Sonification is a subset of the algorithms that translates data from one nature to another one. In our specific case, sonification means a conversion from some kind of data into sounds. In these terms, we can find many types of sonification, such as text-to-speech programs (converting text into audible speech), color readers (color into synthetic voice), Geiger counters (radioactivity into clicks), acoustic radars or MIDI synthesizers. Sonification is, hence, a translation between two sets of data essentially different. Moreover, *translation* should be, thus, replaced by *interpretation*. We cannot guarantee, because of that, that there is no loose of information in this process, since bandwidths are not always (never, indeed) the same. For example, "VGA image, displayed at 25 frames per second, the number of bits per second is: 640 x 480 x 24 x 25 = 184320000 bits per second. The bandwidth of audible sounds is about 15 kHz, allowing a bit rate of 30000 bits per second, so the information has to be reduced by a factor of 6144!" [1].

The example used by Capp is especially worthy for our work, since this kind of translation is exactly what we want to perform, following the guidelines of chapter 4. The proposed system converts images (2.5D gray scale images, as shown in chapter 5) into audible and structured sounds. The conversion from images into sounds will be analyzed and discussed in this chapter.

The outline of this chapter is the following:

- Sound and human hearing system analysis
- State of the art
- Proposal (definition, implementations and evaluation)
- Evaluation
- Discussion
- Conclusions

6.1. The Sound and Human Hearing System

The sounds are mechanical and scalar three dimensional waves. These waves travel through material substrates and can be perceived by the mechanical vibration of some element.

As mechanical waves, they can be described as functions in the time domain or a set of harmonic components in the frequency domain, by means of the Fourier transform (FT):

$$F(\omega) = \int_{-\infty}^{\infty} f(t)e^{-j\omega t}dt \tag{6.1}$$

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(\omega)e^{j\omega t}d\omega \tag{6.2}$$

Equation 6.1 transforms the temporal function *f(t)* in a frequency represented version of the same function, *F(ω)*, being *ω* the angular velocity [rad/s]. The frequency domain functions *F(ω)* can be transformed, again, into its time domain version, by means of equation 6.2. In both equations *j* represents the imaginary constant. Note that the frequency domain representation is a complex function.

The complex sinusoids, as seen in eq. 6.1, work as an orthonormal base for every possible unidimensional wave. These basic real functions have the following structure:

$$f(t) = A \cdot \sin(2\pi \mathrm{f} \mathrm{t} + \phi) = A \cdot \sin(\omega \mathrm{t} + \phi) \tag{6.3}$$

where *t* is the time variable, *A* is the amplitude of the wave, *f* and *ω* the frequency and angular velocity respectively and *ϕ* the phase, in radians. Thus, three parameters must be set for the basic function. A set of these functions, regarding the FT, can generate any waveform.

Another consequence of this formalization is that a sum of different sinusoids is perceived as uncorrelated sound, i.e., noise.

An example of these two equivalent ways of representing the sound is given in the following figures. Only the module of the complex frequency domain function is represented:

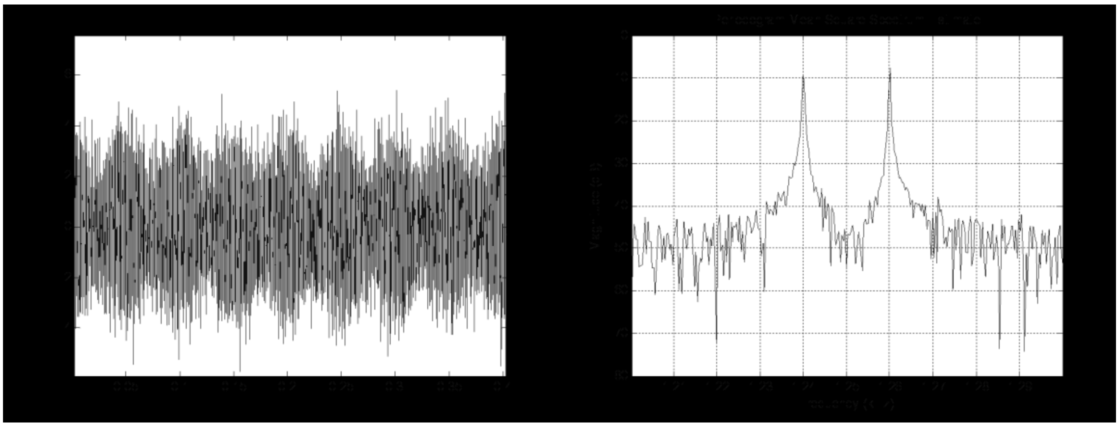


Fig. 6.1. (a) Time domain wave representing two sine waves with noise. (b) Frequency domain representation of the same wave.

In some cases, frequency or spectral domain representation helps to describe parameters or behaviors in the human hearing system and, hence, will be used quite often.

These mechanical waves interact with humans by means of the human hearing system.

The human hearing system is a very complex organization to perceived sounds and must be analyzed from different points of view, since it presents specific properties in terms of biology, psychology, culture, etc.

We can divide it, from a functional and even ontological point of view, in two systems: the biological infrastructure, and the psychoacoustic interpretation. These two essentially different systems must be analyzed separately, since the ways and approaches needed to understand them are completely different. Moreover, even the units and dimensions used in each system are different, the first one dealing with physical magnitudes and the second one with psychological ones.

The following figure presents a functional scheme of the hearing system with the proposed division.

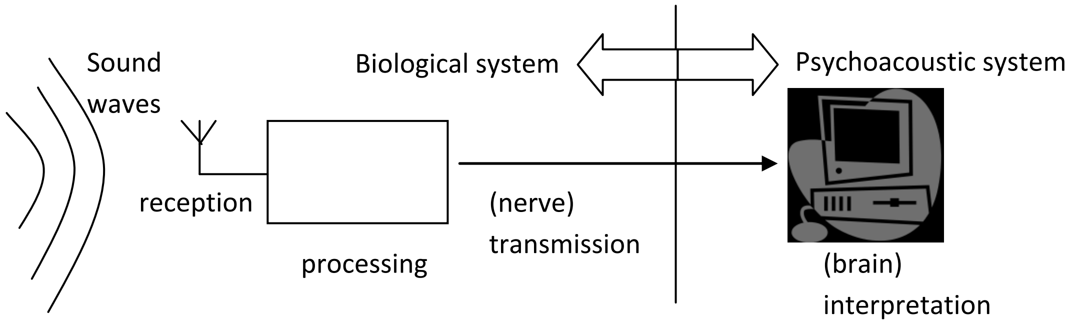


Fig. 6.2. Functional architecture of the human hearing system.

6.1.1. Biological system

From a biological point of view, the system that allows us to hear is composed by several parts. In the following picture one can easily differentiate the different components of the ear.

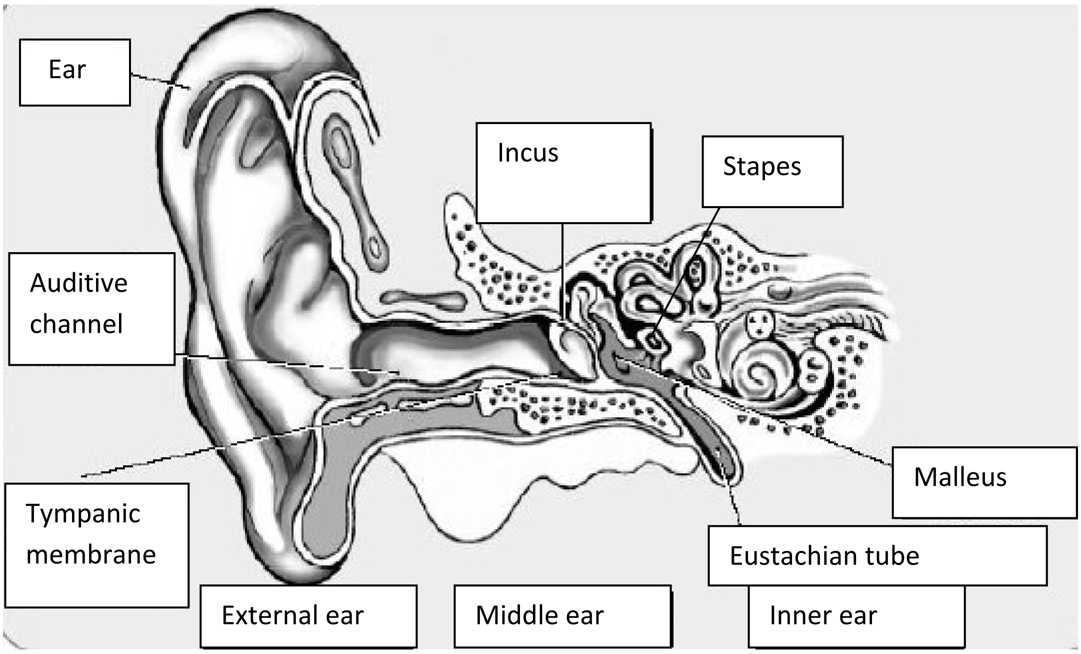


Fig. 6.3. Biological subsystem, from [2], translated into English.

The sound is perceived as follows: The mechanical waves enter in the canal and hit the eardrum. This movement is amplified by means of three bones, the hammer, the stirrup and the anvil, and enters in the inner ear. The function of the three bones is an impedance adaptation between the outer ear and the inner ear. They mechanically convert the low pressure and high amplitude vibration generated in the eardrum, to high pressure and low amplitude vibration in the oval window [3]. This scheme is shown in the following figure.

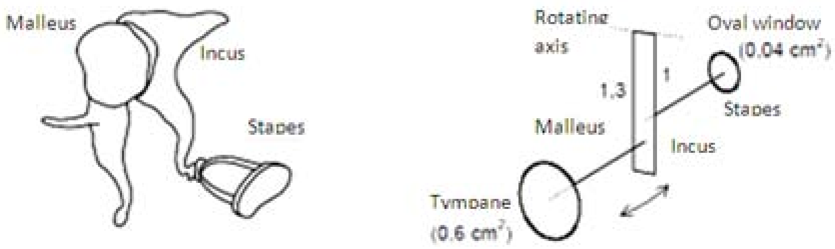


Fig. 6.4. Mechanical impedance adaptation in the medium ear [3], translated into English

Moreover, there are some muscles in this part of the ear, which support the bone chain and protect the hearing system from dangerous and intense sounds. However, this protection is only effective for sound that last more than 500ms [3].

In the inner ear the mechanical waves are converted to nerve stimulations. This process is implemented in the cochlea, which is shown in figure 6.5.

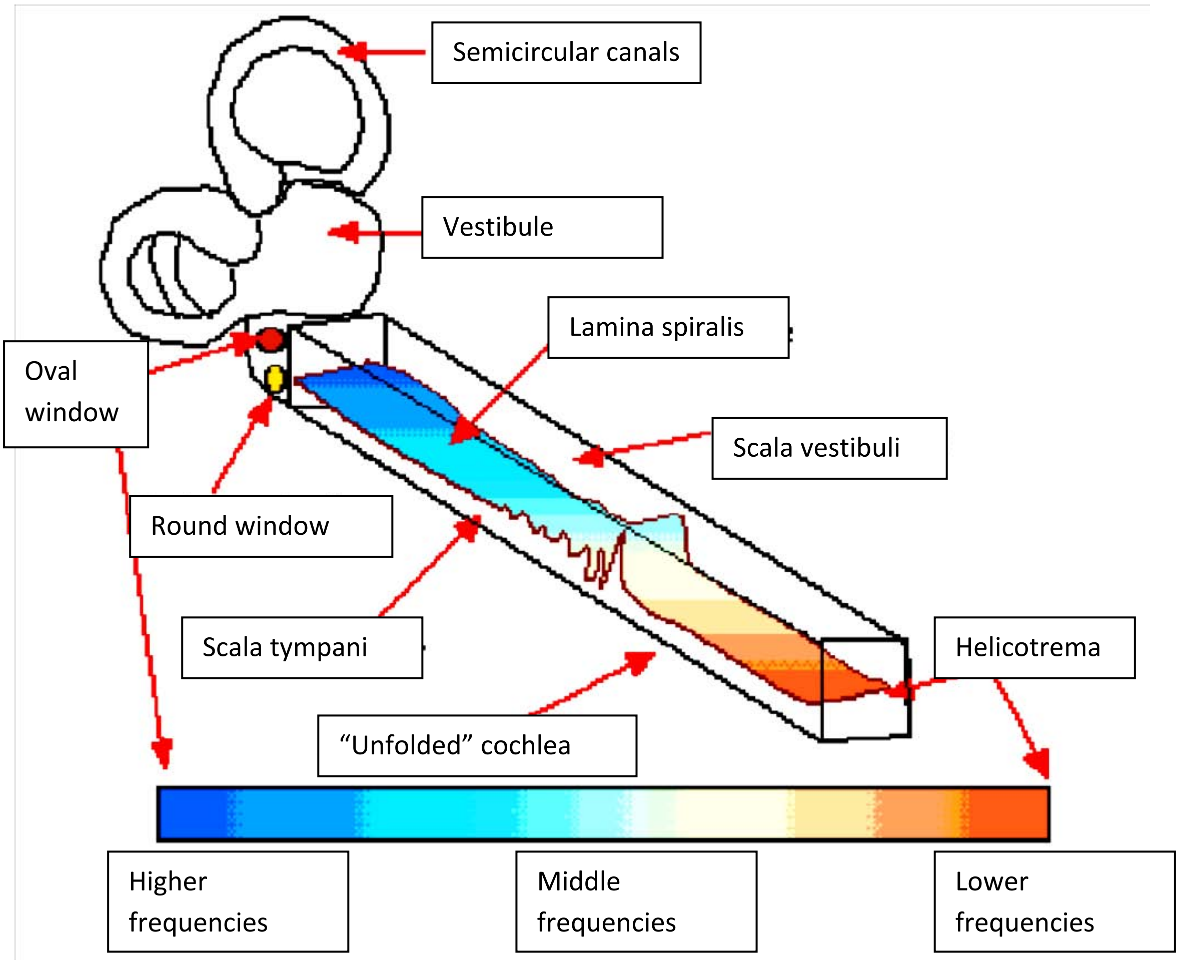


Fig. 6.5. Inner ear, from [4], translated into English.

The cochlea is embedded in the snail bone, and acts as a frequency-to-spatial filter: The cochlea is a closed and, hence, resonant cave, where the mechanical waves reach a maximum at different distances from the helichotrema. The stationary wave stimulates the nerves connected at such distances in the basilar membrane. The brain receives stimuli from different nerves and, thus, frequencies:

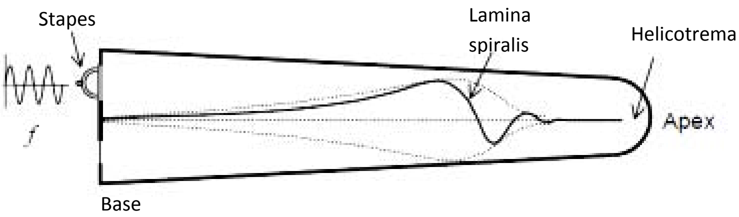


Fig. 6.6. Rectified lateral view of the cochlea with a stationary wave [3].

This system allows the brain to recognize different frequencies, depending on the nerves that are being stimulated by these stationary waves, in function of the distance:

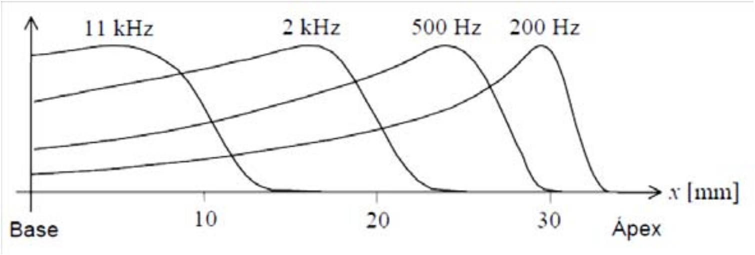


Fig. 6.7. Frequency measurement on terms of the distance to the oval window (base) [3].

6.1.2. Psychoacoustics

The psychoacoustics is the branch of physiology and psychology that studies the psychological perception of the sound.

Once the sounds arrive to the brain, many unexpected effects appear. The sound itself (the physical measure of the waves) is not as we hear it.

There are two concepts that are widely used in psychoacoustic measures [5]:

- Differential threshold: Minimum variation of the analyzed magnitude to perceive a differenced sensation.
- Absolute threshold: The value of the analyzed magnitude which is the borderline between perception and absence of perception.

With these two concepts, the psychological and psychophysical response to many parameters can be explored and the hearing system can be parameterized.

In this line, the first description of the hearing system was proposed by Weber in 1834 [6], stipulating that the "minimum perceived difference" (MPD) is proportional (K) to the intensity of the stimulus (S):

$$MPD=K\cdot S \tag{6.4}$$

This law describes the compression effect measured in every psychoacoustic experiment, although it does not work properly outside the central values of each parameter.

Likewise, Fechner postulated in 1860 [7] that the MPD subjectively correspond to a constant increment of the sensation (S_e) provoked by the stimulus and, thus,

$$S_e = C \cdot ln\left(\frac{S}{S_0}\right) \tag{6.5}$$

where S_0 is the reference stimulus (usually set as the absolute threshold). This expression is the so called Weber-Fechner law.

There is a set of new magnitudes and units that must be defined to deal with the psychological perception of the sound.

The main magnitude to described and identified the sound are loudness and pitch, although we will analyze, as well, other combined characteristics.

6.1.2.1. Loudness

This attribute, defined in ISO 532 and 532B [8] is linked to the intensity of the sound, working in a more complex way, since it depends also of the frequency, the spectral properties of the sound, and others.

The usual way this variable is defined is as sound pressure level:

$$L_p = dB_{SPL} = 20 \cdot log\left(\frac{P_{ef}}{P_{ref}}\right) = 10 \cdot log\left(\frac{I_{ef}}{I_{ref}}\right) \tag{6.6}$$

Where P_{ref}=20µPa (which is approximately the absolute hearing threshold at 1KHz) [3].

Thus, equal sonority level curves can be defined for each frequency (determined by Fletcher and Munson in 1933 [9]:

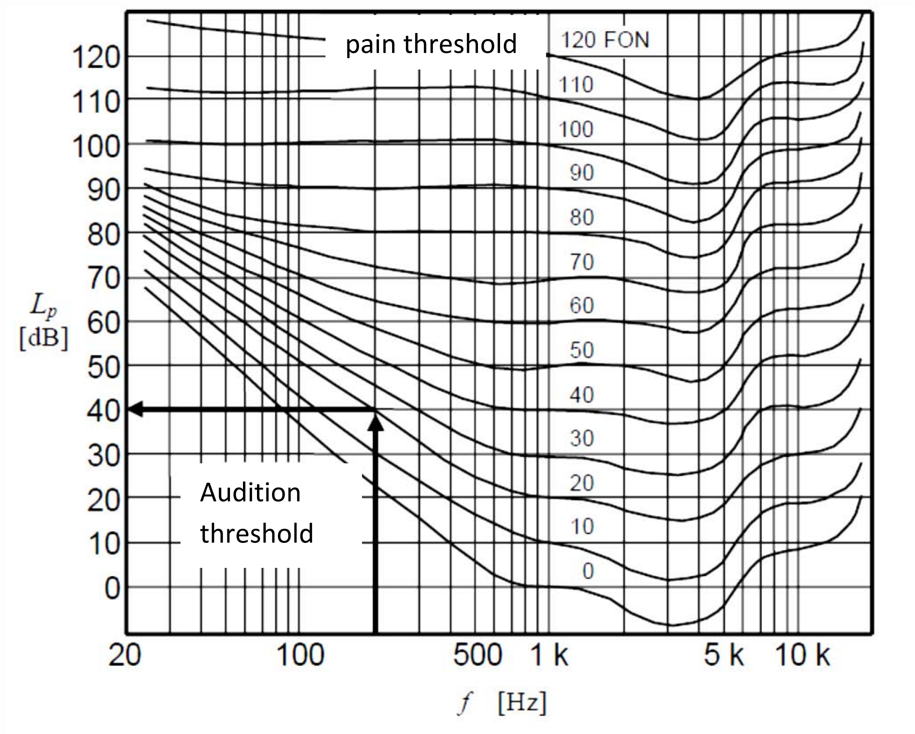
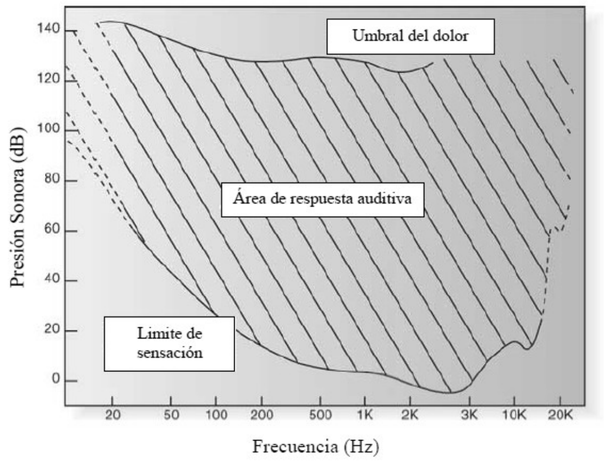


Fig. 6.8. Equal sonority levels [9].

These curves have been standardized by the ISO in the document [10].

In relation with the sonority level of sounds at 1KHz, we can also define the *phon*, setting as reference the audition threshold as 0 phons. 120 phons represent the pain threshold:



Source	P_{ef} [Pa]	L_P [dB]
Pain threshold	20	120
Pneumatic hammer @2m	6.3	105
Piano @1m	0.2	80
Conversation	0.02	60
Room (day)	0.002	40
Recording studio	0.0002	20
Audition threshold	0.00002	0

Fig. 6.9. Hearing range in terms of the frequency [3]. **Table 6.1. Common life SPL references [3].**

The range where hearing is most sensitive the hearing system is between 1000-5000 Hz [11].

The problem with the phones is that they are not perceived as proportional with the psychological perception of the hearing system. Let's remember that we still are dealing with a psychophysic magnitude. The unit that measures linearly the perception of the amplitude is the *son*. The relation between son and phon is shown in the following figure.

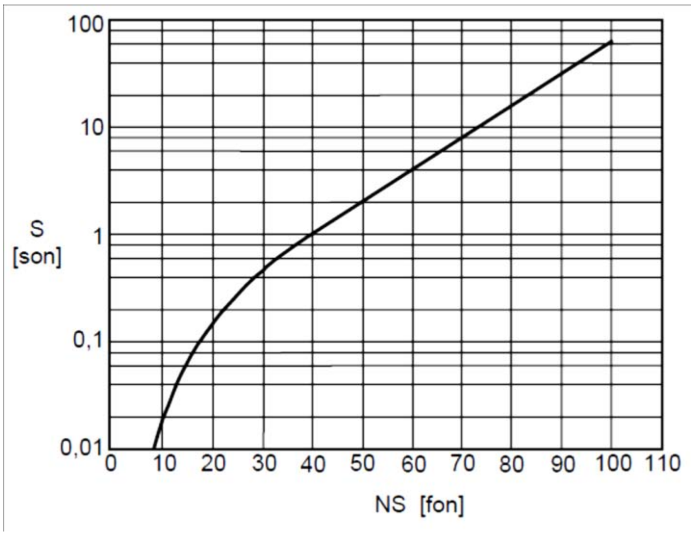


Fig. 6.10. Son VS phon [3].

6.1.2.2. Pitch

The pitch of a sound is a psychoacoustic magnitude related to the frequency, but also with the time of the stimulus and the level, among others. It has been proposed that the pitch is perceived as a log-linear function [12].

The pitch can be measured in two different ways: by means of differential thresholds and with a proportional measurement.

In the first case, the pitch is defined as follows:

$$A(f) = \begin{cases} K_1, f \leq 1000Hz \\ K_2 ln\left(\frac{f}{f_0}\right), f > 1000Hz \end{cases} \tag{6.7}$$

The constant are set by means of a new unit, the *mel*, defined as follows:

$$K_1 = 1\ mel/Hz$$

$$K_2 = 1000\ mel \tag{6.8}$$

$$f_0 = 370\ Hz$$

The relation between frequency and pitch in mels is given in the next figure:

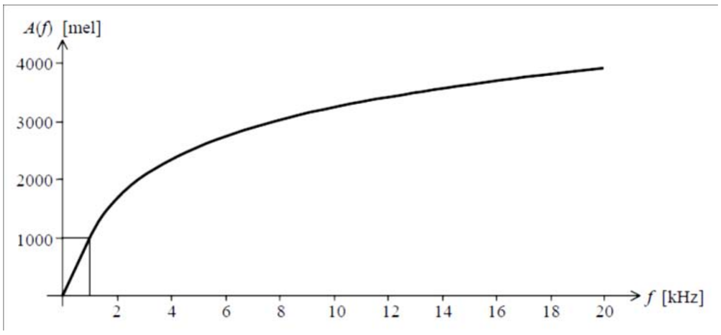


Fig. 6.11. Pitch VS frequency, by means of the differential threshold method [3].

Another option is to ask the subjects to find two tones, the second one the double of the frequency of the first one. This method gives slightly different results:

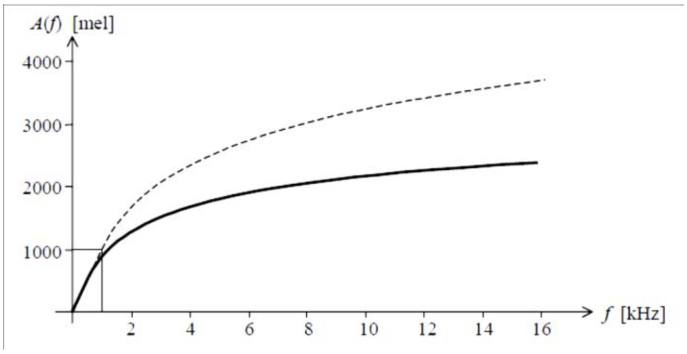


Fig. 6.12. Pitch VS frequency, by means of the proportional method [3].

These scales have reinforced the so called “Theory of the place”, which propose that the psychological perception of the frequency is related with the spatial place in the basilar membrane of the stationary wave maximum:

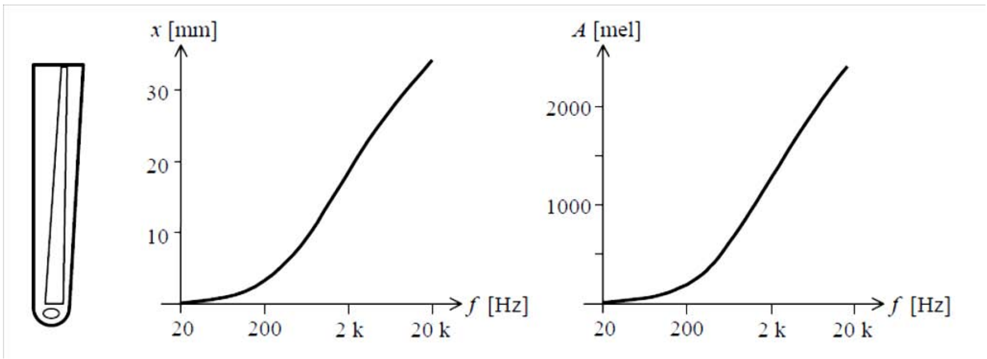


Fig. 6.13. Similarities between the membrane excitation point and the pitch [3].

It is usual to find another unit, when talking about the hearing frequency discrimination, which comes from this comparison with the basilar membrane: the *bark*. One bark is defined as 1.3 mm in the basilar membrane.

Finally, there is an interesting pitch scale, the harmonic one, which comes from the occidental chromatic musical scale. This scale follows a perfect weberian law, with a minimum semantic interval called *semitone*:

$$1st = 2^{1/12} \cong 1.05946 \tag{6.9}$$

Thus, every *octave* is divided in 12 semitones. For every frequency:

$$A_S(f) = 12\frac{\ln\left(\frac{f}{f_0}\right)}{ln2} \tag{6.10}$$

This scale is represented, in terms of the frequency, in mels and semitones, in the next figure:

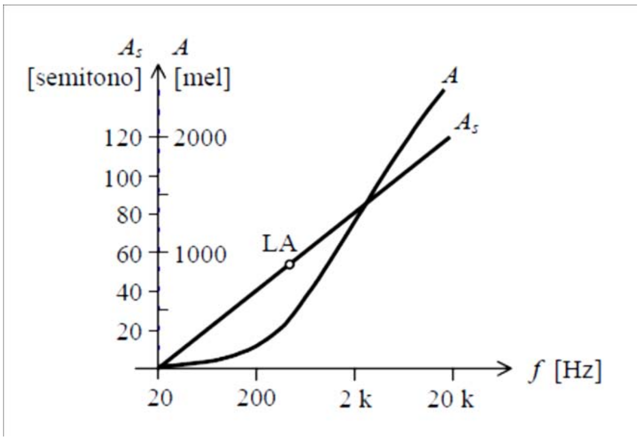


Fig. 6.14. Pitch in semitones and mels. The standard LA note (440 Hz) is marked with a circle [3].

We can see, in this figure, how the pitch measured in semitones is completely linear regarding the frequency.

In the next table, they are shown all the semitones for 6 octaves:

NOTE (EU notation)	NOTE (INT. Notation)	Octave 2 [Hz]	Octave 3 [Hz]	Octave 4 (central)[Hz]	Octave 5 [Hz]	Octave 6 [Hz]	Octave 7 [Hz]
DO	C	65,41	130,81	261,63	523,25	1046,5	2093
DO#	C#	69,3	138,59	277,18	554,36	1108,73	2217,46
RE	D	73,42	146,83	293,66	587,33	1174,66	2349,32
RE#	D#	77,78	155,56	311,13	622,25	1244,51	2489,02
MI	E	82,41	164,81	329,63	659,26	1318,51	2637,02
FA	F	87,31	174,61	349,23	698,45	1396,91	2793,83
FA#	F#	92,5	185	369,99	739,99	1479,98	2959,96
SOL	G	98	196	392	783,99	1567,98	3135,96
SOL#	G#	103,83	207,65	415,3	830,61	1661,22	3322,44
LA	A	110	220	440	880	1760	3520
LA#	A#	116,54	233,08	466,16	932,33	1864,66	3729,31
SI	B	123,47	246,94	493,88	987,77	1975,53	3951,07

Table 6.2. Frequency of each note, in 6 octaves [3].

In a keyboard and a pentagram, these octaves and notes are placed as shown in the following figure:

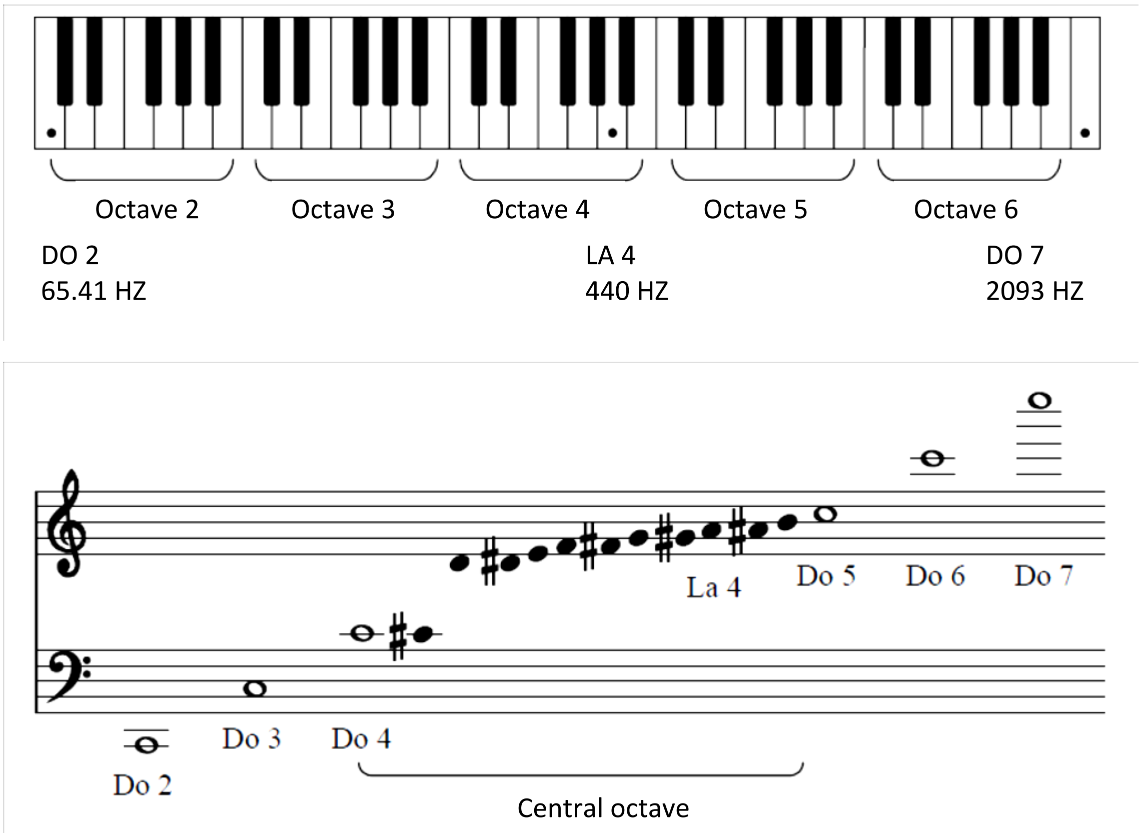


Fig. 6.15. Keyboard and pentagram representation of the 6 octaves and the central LA [3].

6.1.2.3. *Other monaural parameters*

Likewise, the hearing system presents some other effects in the perception of the sound.

On the previous paragraphs, we have been studying the effects of one single tone, and how it is perceived. This is an extremely rare situation in the real world. Every sound is composed of several tones (infinite, indeed, following the Fourier transform). Thus, we have to analyze the response of the hearing system to different combinations of sounds. As a first approximation to the problem, given two tones with different angular frequency ω_1 and ω_2, the superposition of them gives a combined wave:

$$A_1 \cdot sin\omega_1 t + A_2 \cdot sin\omega_2 t = \sqrt{A_1^2 + A_2^2 + 2A_1A_2cos\Delta\omega t} \cdot sin\left(\omega t + \varphi(t)\right)$$

$$\Delta\omega = \omega_2 - \omega_1$$

$$\omega = \frac{\omega_2+\omega_1}{2} \tag{6.11}$$

$$\varphi(t) = arctg\left[\frac{A_1-A_2}{A_1+A_2} tg\left(\frac{\Delta\omega}{2} t\right)\right].$$

Depending on the value of Δω, we can find three cases:

- $\Delta\omega << \omega_1, \omega_2$: If the tones are almost the similar, the hearing system cannot discriminate the difference, since the basilar membrane vibrates in almost the same point:

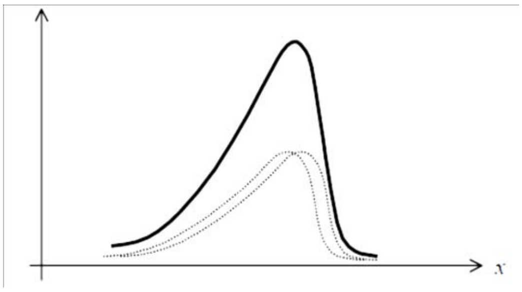


Fig. 6.16. Vibration curve (amplitude VS position) of the basilar membrane for two similar tones[3].

However, we find also a pulsing low frequency tone of value Δω:

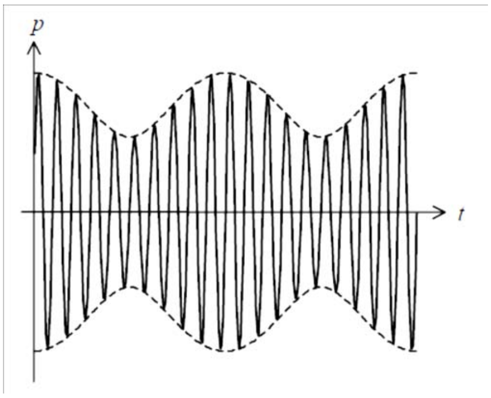


Fig. 6.17. Superposition of two similar tones [3].

- $\Delta\omega > \Delta\omega_d$: The tones may be separated more than the $\Delta\omega_{d,}$ so-called "frequency discrimination threshold". The difference in the pitch can be perceived, but with a dissonance sensation. Moreover, the amplitude of each tone is perceived different even when both present the same amplitude:

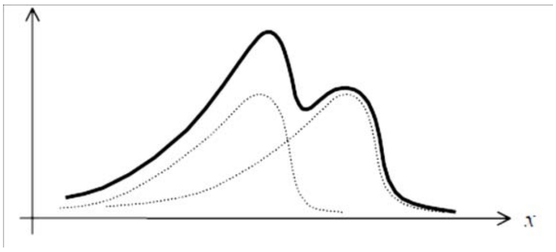


Fig. 6.18. Two slightly different tones. The lower one seems to be stronger than the highest one [3].

- $\Delta\omega > \Delta\omega_c$: Over the discrimination threshold, we find another important parameter of the hearing system, the "critical band" $\Delta\omega_c$. Over this limit, we perceive both tones clearly and completely separated.

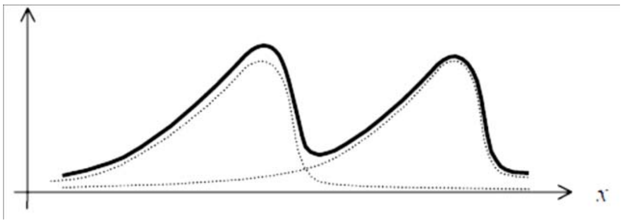


Fig. 6.19. Two different tones[3].

The discrimination threshold and the critical band can be measured and described in terms of the frequency:

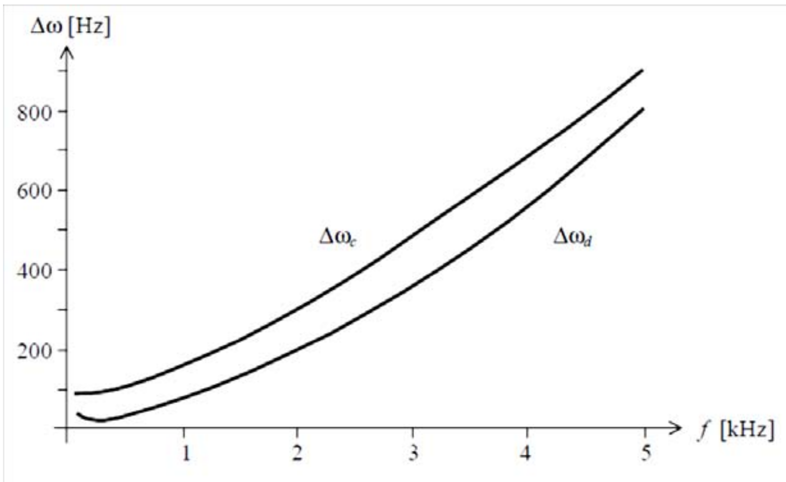


Fig. 6.20. The Critical band and the frequency discrimination threshold in terms of the frequency [3].

These effects express some non-linear responses of the hearing system. One of the most important, related with the critical bands, is the masking effect.

The masking between two sounds can happen simultaneously or sequentially:

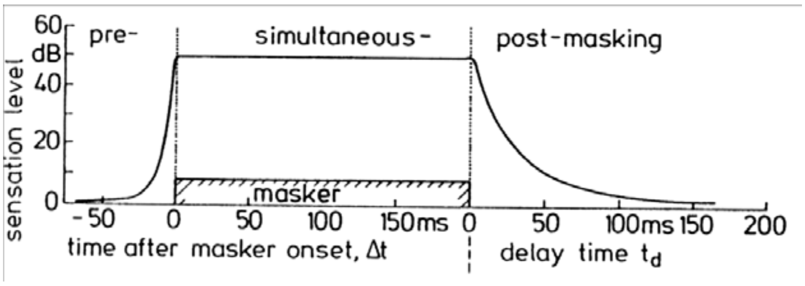


Fig. 6.21. Pre-, simultaneous- and post-masking [5].

The simultaneous masking is the occlusion of one tone by another sound (noise or tone) with different amplitude. If we take a 1 KHz tone with different amplitudes, we can see how they describe a shadow in neighbor frequencies under which no other tone can be perceived. Moreover, this response is non-linear neither with the frequency nor with the amplitude of the test tone.

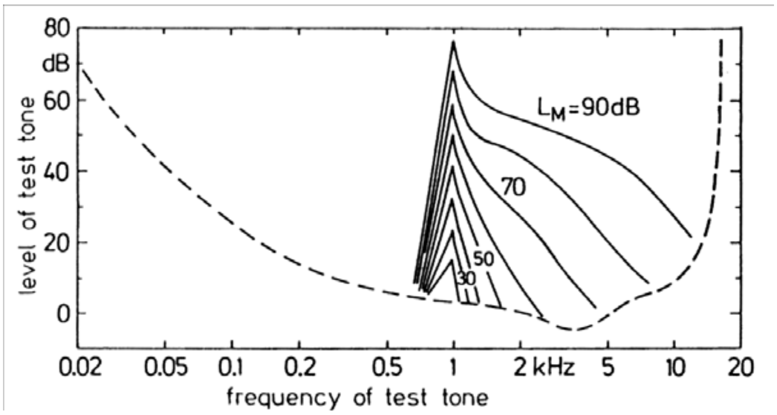


Fig. 6.22. Simultaneous masking of a tone by another test tone [5].

The temporal masking is the effect of shadowing of a tone or a noise that is spread in the time, before and after this reference occurs. The following figure shows how the post-masking works:

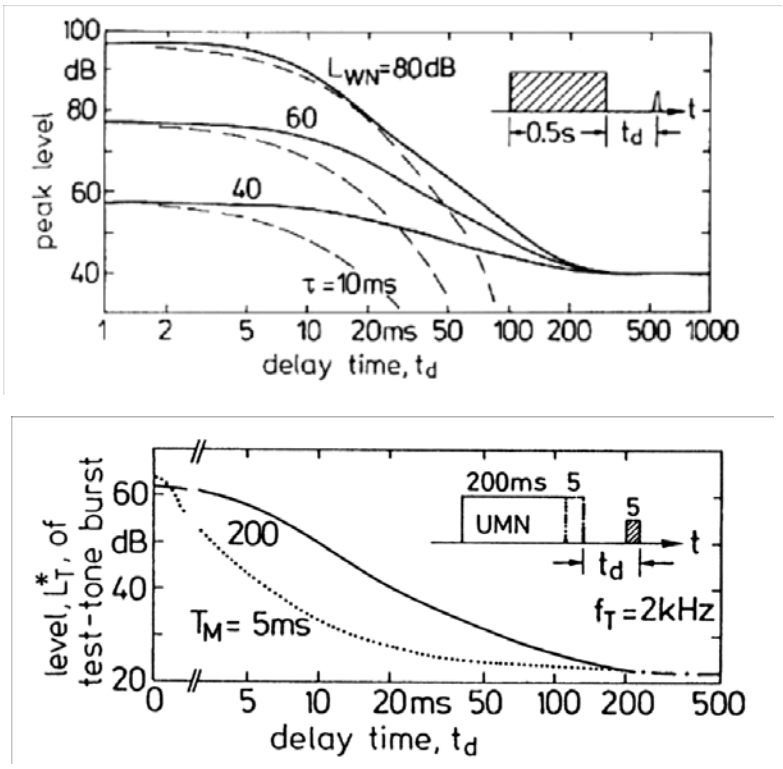


Fig. 6.23. Post masking, (up) the reference tone lasts 0.5s. (down) The reference tone lasts 200ms [5].

We can see a relation between the duration of the masking tone and the masking effect. The shorter is the masking tone, the shorter is the masking effect it provokes.

Regarding the pre-masking, it seems to be an anti-causal effect, since the effect precedes the cause. The masking effect of a sound is spread before the sound even appears. This is a poorly known effect of the hearing system. However, some psychological explanations have been proposed. For example, the sensation requires some time to be formed, and thus it is not instant. Moreover, the time that a sensation needs to be perceived has shown to depend of the magnitude of the stimulus. Hence, a small stimulus temporally precedent and close enough to a bigger one can be shadowed and, finally, not perceived.

6.1.2.4. Distance Perception

Kapralos [13] proposes 5 features of the sound source that allow source positioning:

I. Intensity (sound level) of the sound waves emitted by the source.

2. Reverberation (ratio of direct-to-reverberant sound levels reaching the listener).

3. Frequency spectrum of the sound waves emitted by the sound source.

4. Binaural differences (e.g. ITD and ILD) (see the next subsection).

5. Type of stimulus used (e.g. familiarity with the sound source).

Among them, the first one is the most important. It can be clearly understood this effect from the wave propagation characteristics of the sound as mechanical wave. Moreover, loudness interacts with the reverberation, to allow a deeper comprehension of the distance of a sound source and the environmental characteristics.

Thus, we can assume that the loudness perception is linear (in dB scale) with the distance, as stated in [14] and in the following figure:

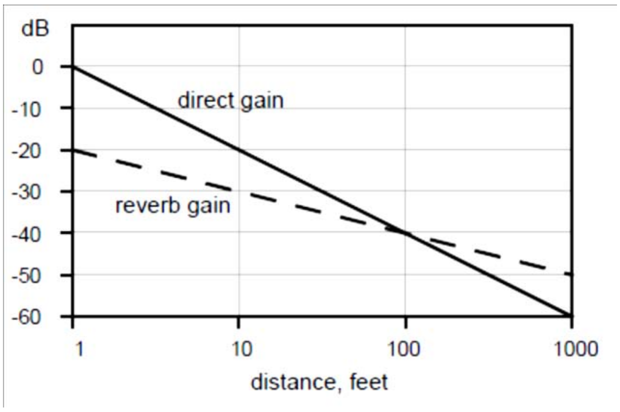


Fig. 6.24. Direct attenuation and reverb attenuation of the perceived loudness of a sound source with the distance [15].

Regarding the rest of parameters, we find in the literature some problems to be solved: Blauert [15] assets that the effect of binaural cues on source distance remains an unresolved issue.

6.1.2.5. Binaural Parameters

Until now, we have been studying the effects of sounds in the hearing system without attending to the binaurality. The human hearing system presents two symmetric systems, which allow perceiving much richer information. Thus, we can say the system is binaural, in opposite to monaural.

The binaurality permits to discriminate the position of the source in the azimuth plane, as proposed Rayleigh in 1907 [16]. Nowadays, we know this process is done by means of three parameters [17]:

- Interaural Level Difference (ILD). This parameter represents the difference of signal intensity between both ears, and it is given in dB. This value also depends of the frequency, since the shadow effect of the head is related to this magnitude. For example, at 500 Hz, the wavelength is 69 cm (4 times the size of an average head). Thus, the ILD is small below the 500 Hz. Finally, the ILD depends of the position of the sound source regarding the head. In the following figure the perception of three sources in terms of the frequency is shown, for an approximation of a spherical head of 15 cm diameter.

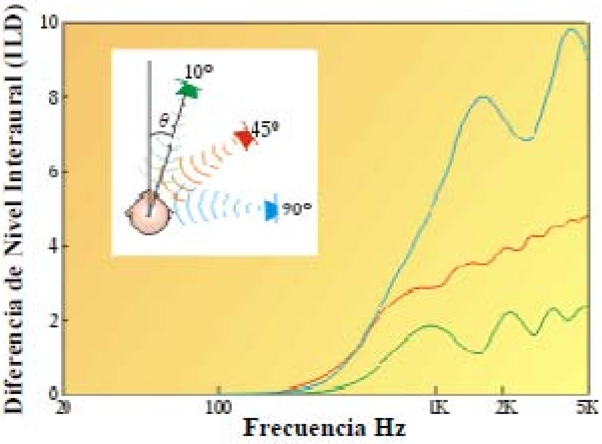


Fig. 6.25. Frequency and position dependency of the perceived ILD [2].

The minimum detectable ILD has shown to be of 0.5dB [18].

- Interaural Time Difference (ITD). Rayleigh proposed in 1907 [16] that the hearing system was able to perceive phase differences between both ears. This difference is related to the delay that is measured between them, by means of the following expression:

$$\Delta t = \frac{\Delta\phi}{2\pi f} \tag{6.12}$$

Taking *r* as the head radius, we can express this dependency as follows:

$$\Delta t = \frac{3r}{c}\sin{(\theta)} \tag{6.13}$$

being *ϑ* the angle of the sound source, and *c* the sound speed (34,400 cm/s). Thus, taking *r* as 8.75cm, *3r/c* is 763µs [19].

The ITD works well for low frequencies, but there is a range where this value is ambiguous, because the wavelength is similar to the head size:

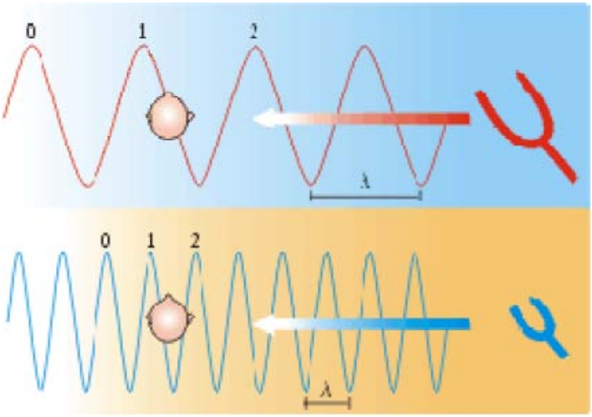


Fig. 6.26. (up) non-ambiguous perception of the ITD. (down) ambiguous perception of the ITD [20].

The range of ambiguous ITD (and, also, ILD) is between 1500 Hz and 2000 Hz. In this range, the hearing system needs a third method to localize the sound sources. Other researchers agree with this statement, as Kyriakakis [21]. However, we find also some research which proposes that ILD is important in all frequency ranges [22], or even in low frequencies, in the proximal region [23].

- Head Related Transfer Function (HRTF). The way the sound is received and propagated inside the head depends of some anatomical aspects such as ears, shoulders, crane structure, neck, torso, etc [24].These elements act as a frequency filter.

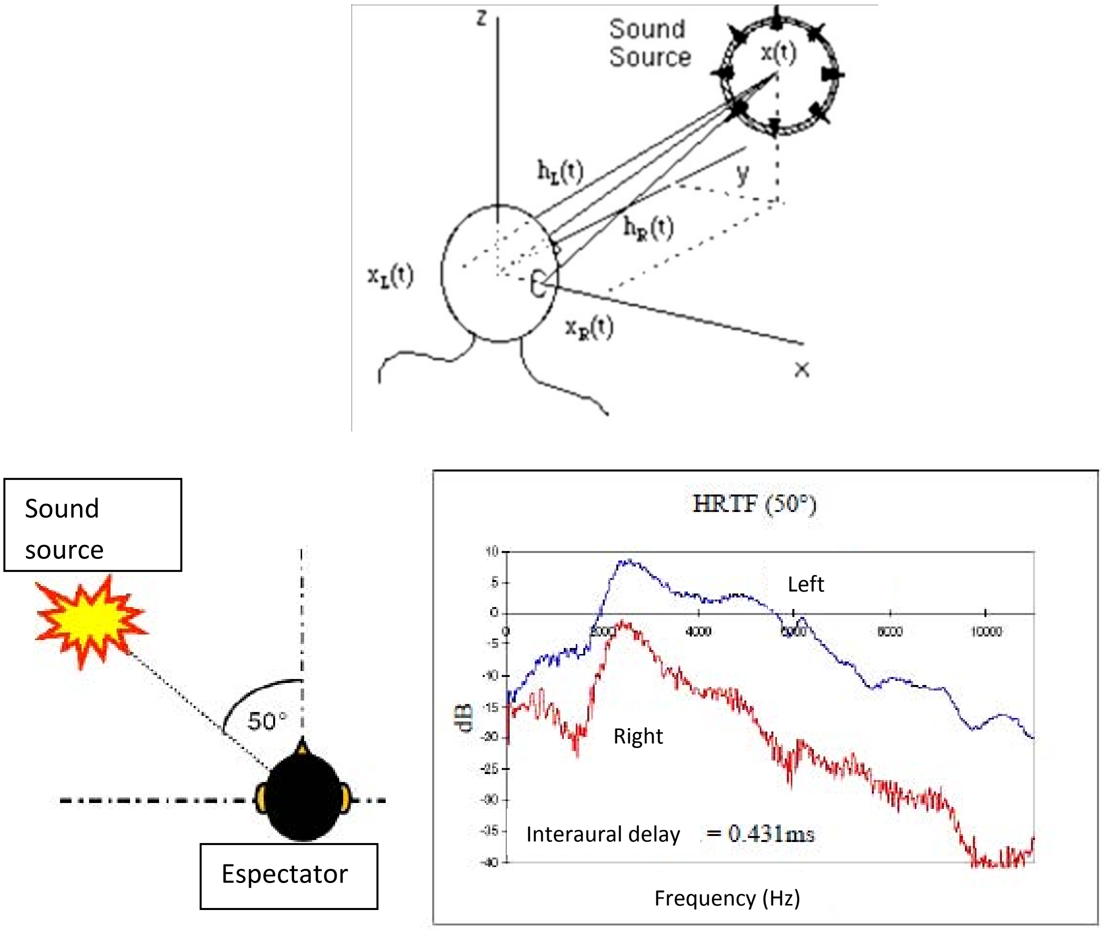


Fig. 6.27. (up) Sound source and head. Different impulsive responses are involved in the HRTF transform. (down left) Lateralization of a sound source. (down right) In blue (left) and red (right) filtering for a lateralized sound source [20], translated into English.

Moreover, sounds coming from the front present an amplification at 1 KHz, while those coming from behind, present that amplification at 3 KHz.

In the azimuth plane, measured errors vary between 5º [20] and 6º [25]. These elements also allow to perceive the elevation of the source, but presenting an error around 9.5º [25].

6.1.2.6. Bone Transmission Specific Aspects

"Bone transmission" or "bone-conduction" (BC from now on) is a possible way of transmitting sounds to the hearing system through the bone structure, specifically the skull and mastoid bones [26].

The BC started to be systematically investigated in the 40's and 50's with studies like [27], [28], [29] and some of them even applied already to the hearing aids systems [30]. However, this field started with preliminary studies many years before [31].

BC is produced by means of different types of mechanical stimulators in some especially relevant points all-around the skull, as shown in the following figure.

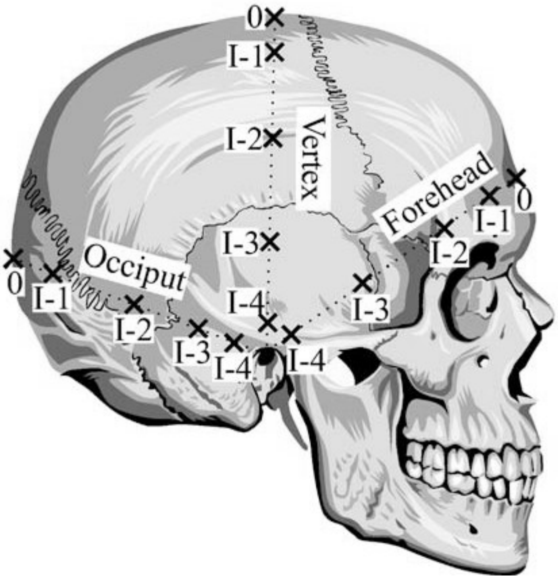


Fig. 6.28. Skull relevant points for the BC [32].

BC is a natural effect produced by mechanical waves penetrating in our body. For example, the difference between our voice as we hear it and our voice recorded and reproduced is caused, as shown in [33] by this effect.

Some psychoacoustic parameters must be redefined of, at least, measured in this different transmission path.

For instance, the skin provokes an attenuation effect which distorts the perceived level through the BC devices. The model and attenuation simulations have been proposed by [34] and it is shown in the following figure.

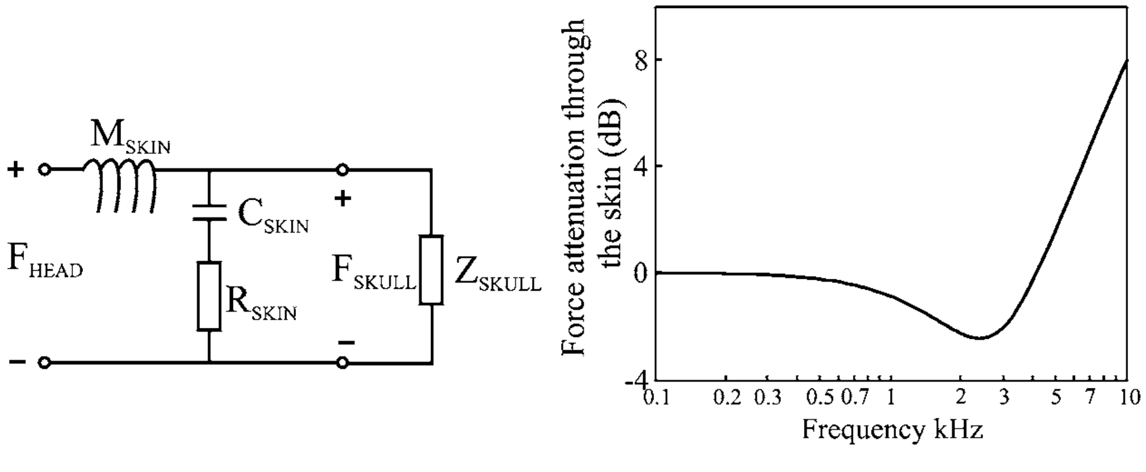


Fig. 6.29. Skull attenuation model and simulated effects.

The parameters of this model were already measured in [32, 35].

Another measurement showed the following impedance response to the BC:

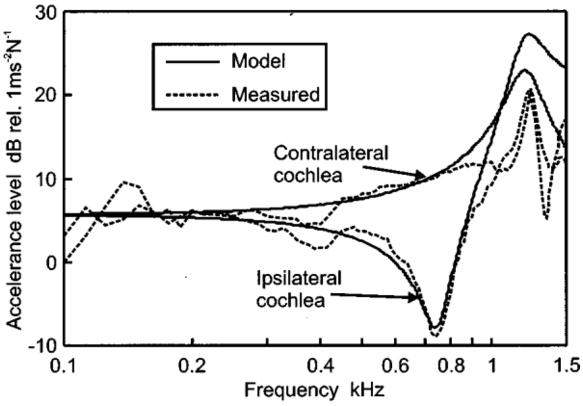


Fig. 6.30. Accelerations measured and modeled in both cochleae. Comparison between a mathematical model and a measured curve [36].

Regarding the figure 6.29, the attenuation over 4 KHz increases dramatically. This has two opposite effects over BC devices. On the one hand, higher attenuations allow discriminating the direction of the source, which is not easy to do when dealing with BC. Regarding the second one, with the frequency range much lower and detailed, we can see where the acceleration is lower, between 0.5 and 0.7 KHz (which provokes an anti-resonance point) and where it is higher, between 1 and 1.5 KHz (with a resonant frequency around 1.4 KHz). The acceleration of the cochleae is inversely proportional to the attenuation. Thus, we can argue that for lower accelerations, the differentiation between left and right will be easier than for higher ones. Thus, interference between both ears is higher in the case of BC than in the air-conducted sound. In that case, the stimulus given to one ear can be completely absent in the other one. However, for the BC something changes. The sound is transmitted through the mastoid from one side to the other one in partial deaf subjects (Mérnière's disease). More in detail, the difference found in [37] are shown in the following figure.

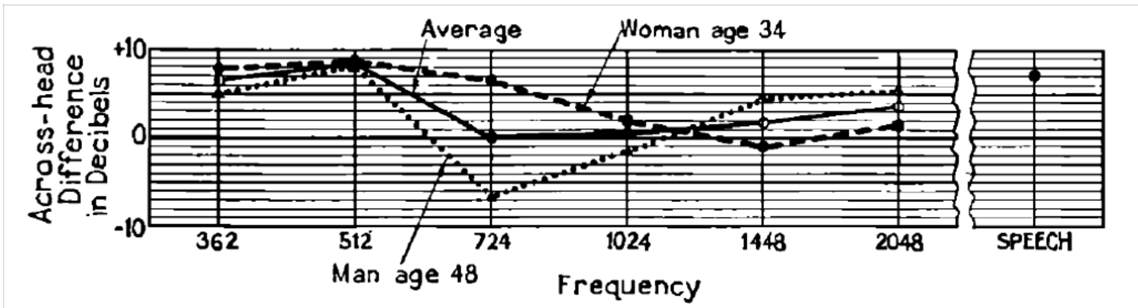


Fig. 6.31. Across-head differences in decibels regarding the frequency. Higher differences are found below 1 KHz and over 1.5 KHz.

This table shows the difference between stimulating the functional side of the hearing system and the stimulation in the other side of the head against the frequency. Only two cases were tested, because of the difficulty of finding subjects with the Mérnière's disease. We can find some limitations, regarding this study, to differentiate the origin of the source when it generates sounds near 1 KHz.

This range (below 1 KHz) will be, then, the easiest to perceive stereo effects.

To avoid the skin effect, researchers have developed the so-called direct bone-transmission, where the bone is directly stimulated by a bone-anchored vibrator [38]. The main problem of this technique is the intrusion needed in the body of the user, by means of chirurgical operations. Thus, this solution will be discarded in our study.

The BC works in both directions, to perceive and to produce sounds by means of vibrations. For example, in [39] there is proposed a microphone attached to a tooth for high noisy environments.

6.1.2.7. Psychological Effects

Finally, we will present two measured psychological effect. The first one, when dealing with real-time sound sources: the reaction time. In [40], an experiment with 16 subjects is proposed, asking them to follow the random pitch change of a tone with a slide control. An example of the obtained results is shown in the next figures:

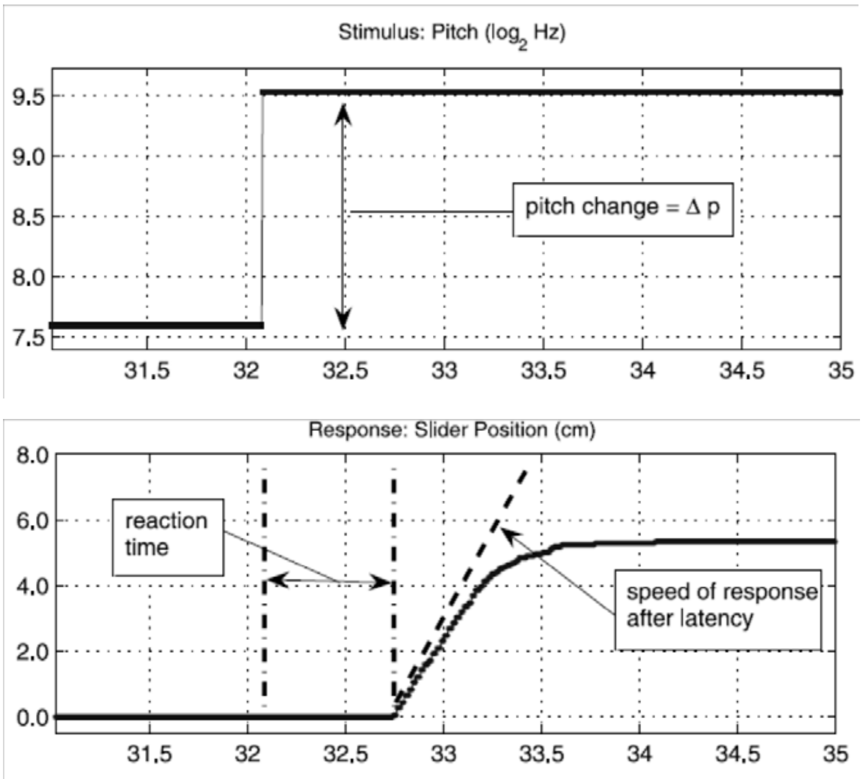


Fig. 6.32. (up) The step function of the pitch. (middle) A participant’s reaction to the stimulus [40].

This will be the worst case. Let’s remark that in this case, a mechanical reaction is demanded and, hence, the reaction time is, obviously, much higher than the psychological perception of the change. This parameter, although is important for the user, it is not taken as relevant for the sonification subsystem. The processing will be implemented with real-time constraints independently of the user’s delay to the change. With that constraint, the reaction will be as fast as the user’s cognitive capabilities permit.

The second one it is a study [25] which demonstrates that blind people do not perceive, as usually thought, more accurately the position of the sound sources than sighted people.

Source type	Average localization error [deg]			
	Sighted volunteers		Blind volunteers	
	Azimuth	Elevation	Azimuth	Elevation
Static – vowel	8.08	15.35	14.42	22.97
Static – wideband	6.74	10.22	12.79	16.75
Moving	6.36	9.47	8.9	16.04

Table 6.3. Error in elevation measurement comparing blind and sighted people, for the "a" vowel and wideband sound sources [25].

Although these discouraging results, we must assume that they are the blinds who need these kind of assistive products and, thus, these additional limitations will have to be taken into account.

An interesting result obtained by Dowling and Harwood [41] states that rhythm aspects overrode their pitch-patterns aspects in determining listeners' responses. This result is accepted by Ramakrishnan [42] to choose rhythm before than timbre as a more salient aspect of music cognition. This result will be discussed in the proposal section.

Finally, there is an interesting study by Yeh *et al.* [43] which states that the brain is able to estimate the fundamental harmonic even when it is not present in the sound, if the rest of harmonics are present.

6.2. State of the Art

The use of sounds to describe images begins with the articulated language itself. For instance, a word (the sound of a word) describes some region of the real (or imaginary) world. Although, we will focus on sonification proposals dealing with non-articulated language, we also find some examples of "verbalization" instead of sonification approaches, especially in the field of the technical aids for the blind.

6.2.1. Historical Approach

The use of sounds to describe the visual world has been proposed since the last years of the XIX century. Specifically, Noiszewski built the Elektroftalm in 1897 [44, 45]. Some years later, in 1912, Albe built the Exploring Optophone [1, 45, 46]. Of course, these two approaches where extremely simple: the Elektroftalm used a selenium cell to discriminate whether there was light or not and, then, producing a sound to inform the user. The optophone, in the same way, produced a sound proportional to the light received by the selenium cell. These first two assistive products received critics from the blind community, because of the useless of such simple information. A more complex proposal came, in those years, also from d'Albe, with the Reading optophone [1]. This new system, allowed perceiving different combinations of sounds, representing characters written in a book, as shows the following figure.

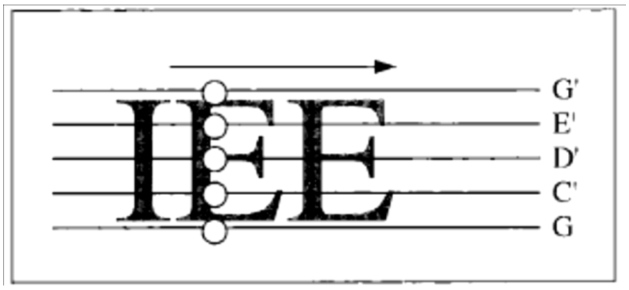


Fig. 6.33. The Albe's Reading optophone, 1913 [1].

> "The idea for this was that a musical 'motif' would be produced from the reading of text, which would be far more pleasing to the ear than a set of unfamiliar sounds. […] One expert student of the device, Miss Mary Jameson […] attained a reading speed of 60 words/min, although the average was generally a lot lower" [1], p. 139].

The Albe's Reading optophone should be recognized as the first Optical Character Reader (OCR).

As a final curiosity, we can retrieve from the 40's decade the Radioactive guider [47], which used a radioactive element to produce a beam of radiations that were reflected in the bodies and detected by a Geiger counter, thus, converted into clicks.

Since the late 70's and early 80's, the technology allowed to design and build more complex, fast and smaller devices.

6.2.2. Review of Sonification Codes

We will call "sonification code" to the rules that make a visual cue correspond to an auditory cue. For example, we can find in [1] the following rule: Loudness is normally associated with brightness, but a more normally representation is to associate loudness with proximity to the user. As another example, the sonification shown in figure 6.33 follows the so called *piano transform*.

In this line, we can classify and gather the sonification proposals found in the literature in terms of the parameters used to acoustically describe an image, as well as the number of channels and the dimensional complexity of the produced sonification.

Following Milios' study [48], we can identify two modes of sonification functions:

- Proportional mode, where the audio feature is a nonlinear function of the instantaneous value of range. More precisely, the nonlinear function is a decaying exponential mapping. In this case, the function follows the so-called weberian law, and will be used widely in one of the paradigms studied herein.
- Derivative mode, where the audio feature is a function of the temporal derivative of range. More precisely, changes between consecutive range measurements (an approximation to mapping the derivative of range as a function of time) are mapped to the audio domain. This option, although has been proposed, as we saw in the Interviews chapter, is not often used. Indeed, we will not find any ETA implementing it. One of the reasons is that our hearing system (and, with some exceptions, every sensory system) works mostly following the first option, i.e. the weberian law. Moreover, our perception does not need temporal changes to perceive information.

Another reason is that the global accuracy if the first option is higher than this second one, as shown in the following figures.

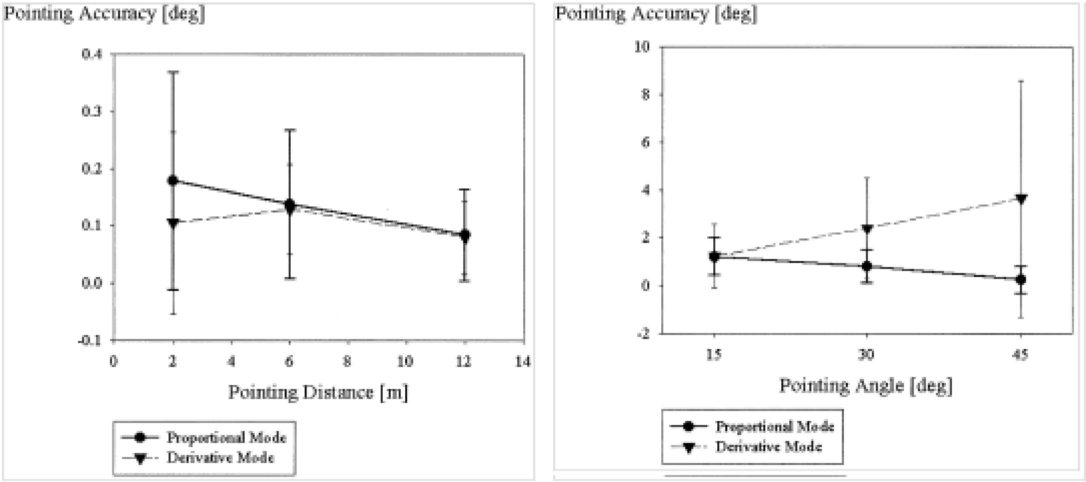


Fig. 6.34. (left) Means and standard deviations of pointing accuracy when pointing while in the proportional (circles) or derivative (triangles) mode at the sound source located at 2, 6, and 12 m. (right) Means and standard deviations of pointing accuracy when pointing while in the proportional (circles) or derivative (triangles) mode at the sound source [48].

We propose, in this work, two different taxonomies in terms of the number of channels and, more in detail, regarding the sonification paradigm.

6.2.2.1. Classification of Sonification ETAs by the Number of Channels

The analyzed sonification electronic travel aids in the chapter 2 can be gathered following the number of channels they use to transmit the information. Only two groups can be formed following this criterion:

- Monaural: One single channel is used to transmit the information. This information uses to be codified in a continuous dimension, such as frequency, amplitude, etc. Thus, most of them are unidimensional. Some examples of this group of ETAs are the Nottingham OD, US Torch, Mims, FOA Laser cane [49] or the Sidewalk detector [50].

- Binaural: Binaurality allows transmitting much richer information to the user. In fact, one more dimension. Thus, some more complex proposals can be found in this family. Some of the most important ones are the Navbelt [51], Multi and cross-modal ETA [52, 53], Echolocation [54], EAV [55], 3-D Support [56-58], CASBliP [59, 60], vOICe [61], Sonic Mobility Aid [62], NAVI [63], Sonic Pathfinder [64-68], Sonic Guide [69, 70], etc.

We can easily see the difference in the amount of proposals found in each option. Binaurality allows a more complex understanding of the environment, thus, it has been much more used for mobility assistive products.

6.2.2.2. Approaches Analysis

In both mono and binaural solutions, some choices must be taken to determine how the ambient information is encoded. Any decision belongs to one of the two possible sonification paradigms:

- Psychoacoustics: This paradigm employs the natural discrimination of the source spatial parameters (distance, azimuth and altitude for instance). It uses the functions and curves described in the first section of this chapter, to simulate, by means of convolution processes, the virtual position of the source.
- Arbitrary: For attributes that are not directly related to sounds (such as color, or texture, for example), or to substitute spatial parameters by some other not naturally code, many ETAs use arbitrary transformations between a physical property and the associated sound.

We will now discuss in depth each paradigm. However, there is almost no "purely neither psychoacoustic" nor "arbitrary" sonification proposal and most of them present a mix between these two paradigms. Just for presentation purpose, we will talk about some clear examples of each case, focusing on the most relevant features regarding these paradigms.

6.2.2.2.1. Psychoacoustics

The psychoacoustic paradigm uses, as said, the natural effects in the hearing system to help localizing the sources positions.

Rossi *et al.* propose a classification of psychoacoustic implementations in 6 levels in their sound immersion level scale [71]:

Level	Techniques/Methods	Perceptions (results)
0	Monoaural "dry" signal	No immersion
1	Reverberating, echoes	Spaciousness ambience
2	Panning (between speakers), stereo, 5.1	Direction movement
3	Amplitud panning, VBAP	Correct positioning in limited regions
4	HRTF, periphony (ambionics, WFS…)	Stable 2D sound fields
5	HRTF, periphony (ambionics, WFS…)	Stable 3D sound fields, accurate distance and localization

Table 6.4. Rossi's acoustical immersion levels.

Every modern ETA implements at least acoustical environments over level 2. The most advanced ones, even in the 4 or 5 levels. Such is the case of the Bujacz's proposed ETA [72]:

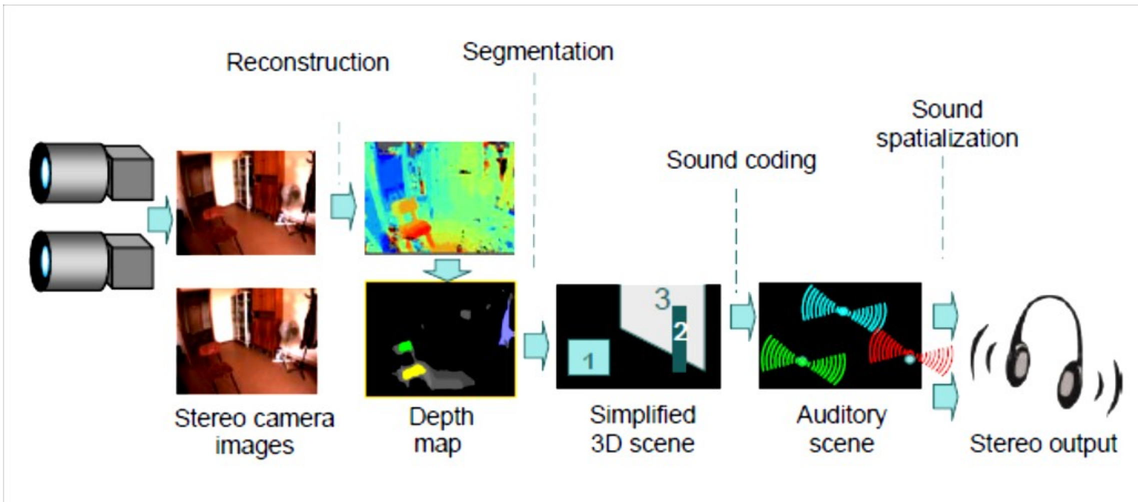


Fig. 6.35. Psychoacoustic sonification based ETA,

In these cases, the user perceives different sounds filtered as if they came from different spatial situations. Thus, these systems implement several convolutions (i.e. discrete sum of products of delayed functions) for each sound source to virtually place them in the correct solid angle and distance.

The main advantage of this proposal is, obviously, the naturally situation of the sound sources. No training is needed at all, and the system can be used by untrained users since the beginning.

For example, regarding the azimuth localization, the next figure shows the intensity mapped in each ear versus the virtual angle of the source:

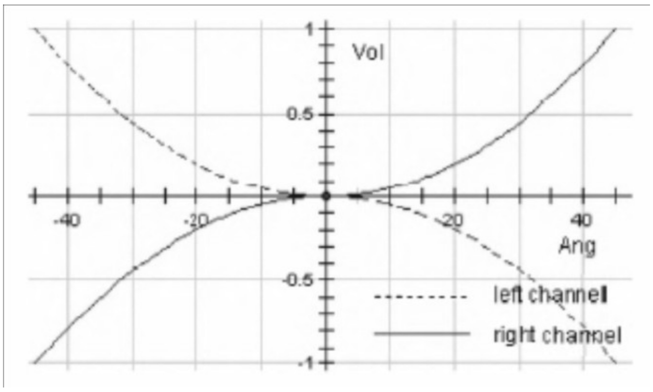


Fig. 6.36. Volume-Horizontal Deviation Degree Mapping Curve [73].

However, some smaller problems stay undetermined, such as “To determine if the most naturally perceived direction information is obtained when (a) referenced to the person’s head orientation, or (b) referenced to the person’s body orientation.” [74].

More problems can be seen after analyzing existing psychoacoustic proposals: in vertical localization, HRTF proposals present high errors, as shown before. Some projects even do not implement such processing, as the already presented CASBliP [20]. In this case, the user must move the head, since the azimuth processing is done in the plane of the eyes.

Likewise, the computation load is high, because of the two complex convolutions needed for each sound source. As Castro Toledo states, “Avoiding the poor perception of the elevation the use of individualized HRTFs functions is recommended. An individualized HRTF function is computationally-complex and cannot be used for real-time spatial rendering of multiple moving sources.” [75].

In the case of distances sonification, the psychoacoustic paradigm is followed by every analyzed project; however, some issues can be pointed out:

- Some of them link, as well as the loudness, the frequency to the distance [72].
- Others, may propose an inverse codification of the distance, however, following the psychoacoustic curve of distance perception, as done in [73]:

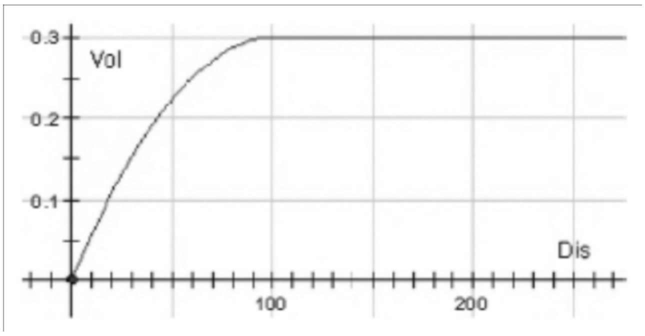


Fig. 6.37. Volume-distance inverse mapping curve of Zhigang's project.

Finally, examples of ETAs mainly implemented following this paradigm are, among others, the Navbelt [51], the Multi and cross-modal ETA [52, 53], the Echolocation [54], the EAV [55], the 3-D Support [56-58], the FIU Project [76], the Remote Guidance System, [77] or the CASBliP [59, 60].

In the case of the EAV, from the Instituto Astrofísico de Canarias and the Laguna University, we find an example of sounds generated by means of this prototype in the following link.

6.2.2.2.2. *Arbitraries: 4 Paradigms.*

There is an interesting distinction done by Walker and Kramer [78] which separates *intentional* sounds as those "purposely engineered to perform as an information display", from the *incidental* sounds "which are non-engineered sounds that occur as a consequence of the normal operation of a system". In many psychoacoustic based sonifications, we can see how clicks and other *incidental* sounds are used to transmit the information (i.e. the information does not travel in the click itself but in the spatialized transform of this click). We find, then, another branch which tries to transmit information in the sound itself. This is what we called *arbitrary* sonification.

In general terms, arbitrary sonification uses some other characteristics of the sound, such as frequency, brightness or timbre, formants, saturation, etc. which, not being related with physical characteristics or parameters of objects or the surrounding, can be linked artificially to them in order to transmit more information. This linking is, then, arbitrary, i.e., there is no physical/mathematical reason why we should follow one option against another one. Thus, we will find in this section several different proposals which will be critically analyzed. Te main problem of this approach is the needed learning. The main advantage, as said, of the psychoacoustic transforms was that no training is needed since the same naturally developed skills are used to localize the virtual sounds. Likewise, in the arbitrary sonification, users must learn how real world objects and sounds are related, which depends, on its turn, of the final paradigm elected.

One of the deepest studies dedicated to this field was done by O'Hea during his Ph.D. research [79]. Following this research, Capp [1] divides the different proposals in 2 paradigms, to which we will add another two:

- Fish [80] used frequency to map vertical position, and binaural loudness difference for horizontal position (we can see here a psychoacoustic implementation for this second dimension). Brightness is mapped to loudness. We do not discuss in this very moment if this brightness is related with the natural brightness of the scene, or with other parameter such as the deep, after a scene processing. All these three parameters change in frame-by-frame and is usually called *point mapping.* The basic process of this transform (for a 8×8 image) is shown in the following figure, modified from that of [61]:

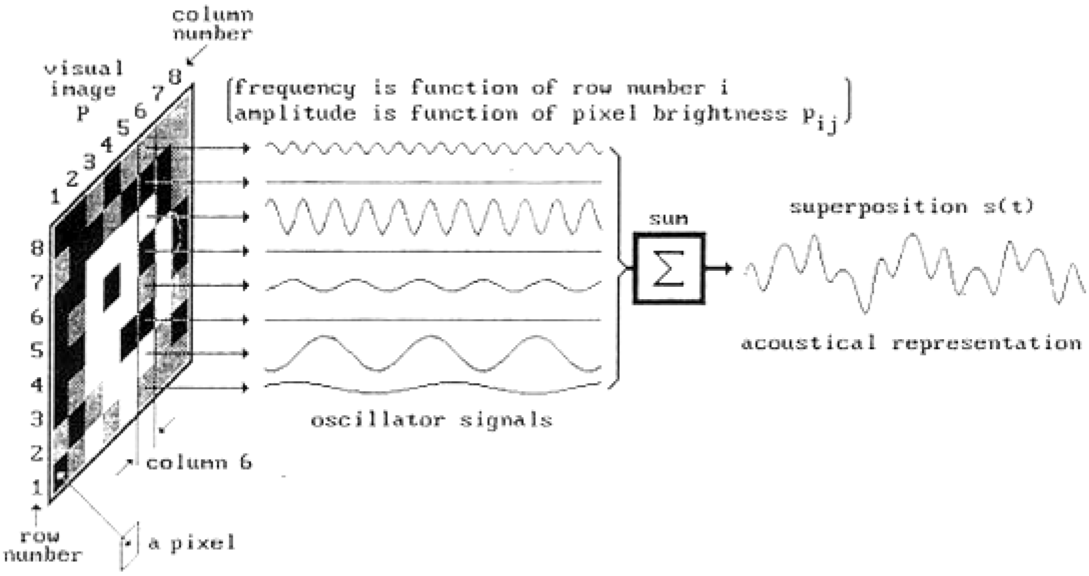


Fig. 6.38. Scheme of the point mapping.

For example, the SVETA project [81] generates the sound following the next expression:

$$S(j) = \sum_{i=1}^{N} I(i,j)M(i,j) \tag{6.14}$$

Where *S(j)* is the sound pattern produced from column j of the image, *j* = 1,2,...,16 and *j* = 32,31,...17 for stereo type scanning, *I(i,j)* is the intensity value of *(i,j)*th element, and *M(i,j)* is the sample of musical tone for *(i,j)*th pixel. In this case, the brightness is related with the depth of the pixel. The main advantage of this transform is the instant representation of the scene. For any static image, a stationary sound pattern is created. The problem is that, since the time is not a function of the scene, the information is embedded in the same pattern and might be more difficult to discriminate small details in the perceived pattern.

- Dallas [82] mapped again vertical position to frequency, horizontal position to time, and brightness to loudness. The mapping used is an example of the *piano transform* (shown in figure 6.27). A modern example of this transform can be found in the VOICe project [61]. In this case, a click signals the beginning of the scene description:

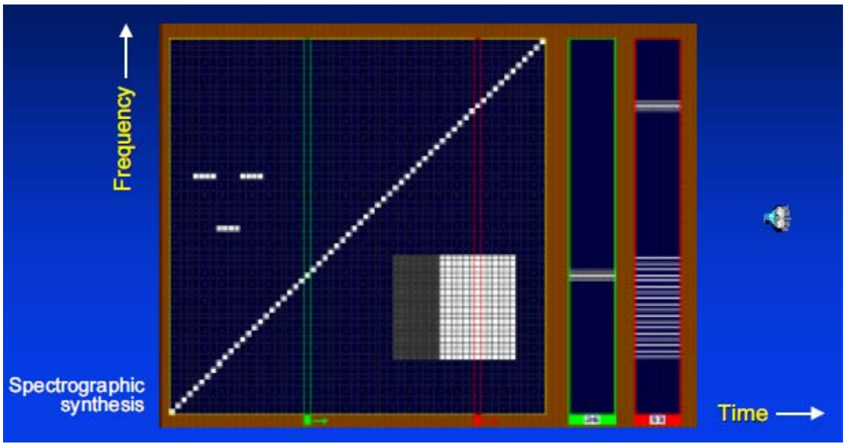


Fig. 6.39. Simple scene in two consecutive moments [83].

As it is done in the first paradigm, there is a version of the VOICe which maps left sounds (i.e., the first ones that are reproduced) to the left ear, and the right ones to the right one, thus, helping the understanding of the scene by means of psychoacoustic transformations. The main advantage of this option is the temporal structure of the scene information, perceiving sequentially the information of the scene from the left to the right. However, we have to point out that this mapping does not allow real-time implementations, since it requires a gap of time to represent the scene. In the case of the vOICe, “the allowed conversion time T is restricted by the fact that the information content of an image soon becomes obsolete within a changing environment. Furthermore, the capacity of the human brain to assimilate and interpret information distributed in time is also limited” ([61], p.115). The time T is set to lay between one and two seconds for 64×64 images. Examples of this sonification proposal can be found in the following figures (click on the figure to hear the sonification if the resource is available in “.\resources\”).

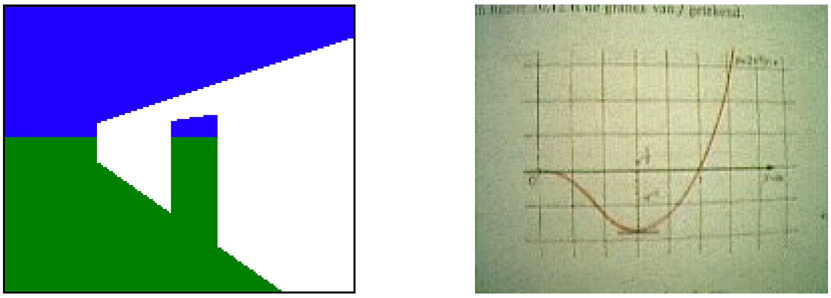


Fig. 6.40. (left) Wall 2D representation plus sound [www.seeingwithsound.com/voiswall.htm] and (right) a graph sonification (from www.seeingwithsound.com/prgraph.html). Resources must be available in “.\resources\”.

There is also available a video of this transform which can be accessed through the following link.

- We can add to these two mappings another one which maps distance to frequency, proposed by Milios [48]: “to be mapped to higher frequencies thereby stressing their importance of objects nearby. The maximum frequency in this mapping (4200 Hz) corresponds to a range measurement of 0.30 m whereas the minimum frequency (106.46 Hz) corresponds to a range measurement of 15 m” (p.

417). We will call this option the *pitch transform*. The main problem of this approach is the learning process for a dimension somewhat intuitive such as the distance, giving and advantage in the case of highly noisy environments, where differences of loudness might not be perceived and differences of pitch, at constant volume, work much better.

- Finally, and forcing the mapping concept, we will propose the *verbal transform*, i.e., traducing the scene into words (synthetic or recorded) which allows the user to form a mental representation of the surrounding. For example, the Mini-radar [84] produces messages like “Stop” or “Free way” depending of the surroundings. The main advantage is that no training is needed in this case (just, if any, the language), losing, on the other hand, a huge amount of information, since these kind of transforms use to recognize simple forms or structures which are attached to specific signals. Thus, complex scenes are hardly convertible automatically to speech.

Additionally, we could ask what should the brightness represent, and here we find, as before, different options:

- Light intensity: This is the first implementation provided by researchers. This is also the case of the first electronic assistive products developed, as we saw with the optophone or the elektroftalm. In the literature, this mapping is called *direct mapping* [81]. This option has been widely used in modern ETAs as well. This is the case of the vOICe, already analyzed. We find here problems already present in the reading optophone. If the image is mostly white or bright, the generated pattern is noisy and can hardly be understood. Moreover, the light intensity is not related at all with the significance of the information. For example, the sky is bright, but completely irrelevant for mobility aspects. Another problem is that it is impossible to estimate the distance of an object, so it hardly serves for mobility purposes.
- Depth: An alternative to represent directly the image captured by cameras into sounds, is to represent the depth. In this case, the closer the objects are, the higher is the sound. This is widely implemented, since the relation of importance with the distance to the user is obvious. The SVETA [81], for example, uses this approach and, specifically, the musical octave sonification of 2.5D images. An important problem of this approach is the loose of plain information, such as text written in a flat surface.
- Edges: To make the system able to recognize information or patterns over flat surfaces, an edge representation has been also proposed. The advantages over the depth map, but also its limitations, can be perceived in the following figures.

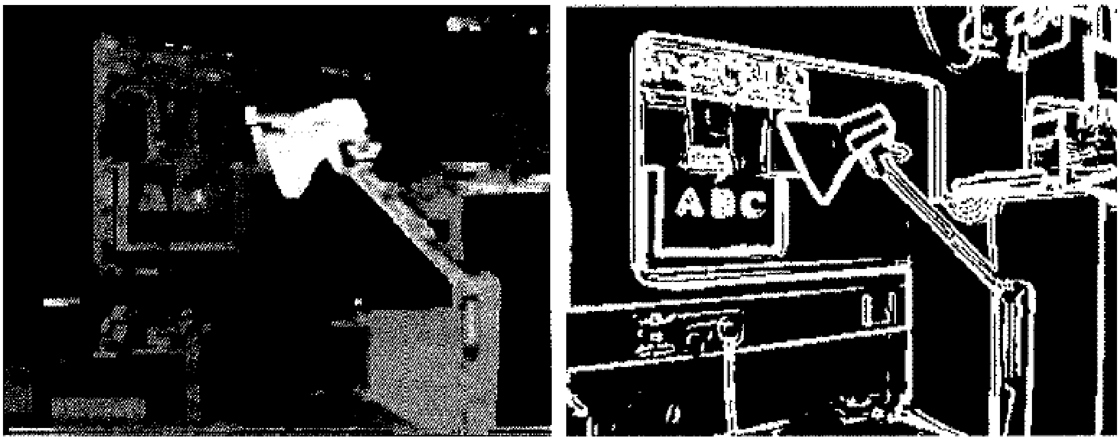


Fig. 6.41. (left) Depth map representation (with edge pre-filtering) and (right) edge representation [1].

There are not too many examples of the direct or sonification approach. Among them, we find the monaural version of the vOICe [61], the proportional sonification proposed by Milios *et al.* [48] or the Sonic Mobility Aid [62].

The main reason is exposed in [85], by means of the following table:

Object	Percentage predicted	
Characteristics	Direct Mapping Method	Musical Octave Method
Position (Top/Bottom)	66%	78%
Position (Left/Right)	100%	100%
Shape	72%	88%
Size	77%	91%
Distance	32%	98%
Pleasantness	40%	91%

Table 6.5. Accuracy determining spatial parameters by means of two sonification methods: direct and musical octave, from [85].

6.2.2.2.3. Mixed Solutions

Although we dedicate a specific section to the mixed solutions in sonification, we have already seen some of them in the arbitrary approach. The mixed approach tries to overcome the problems of both previously presented options:

- On the one hand, these systems use to take from the psychoacoustic approach the binaural capacity of the hearing system, to apply it to azimuth localization of the sound sources. As well, the distance is usually related to the loudness of the sound, following a weberian law.
- On the other one, elevation and other non psychoacoustically-related characteristics (texture, color, edges, velocity…) are mapped into arbitrary parameters, trying to reduce the error of the localization (in the case of elevation) or to represent extra information, which could be relevant.

The main advantage is, mainly, a reduced learning process regarding the strictly arbitrary approaches. The main drawback is that, however, training is then mandatory because of the arbitrary chose of some parameters.

We find in this group a lot of assistive products, such as that presented in [63], the Sonic Pathfinder [68], the Sonic Guide [69, 70], the Single Object Sensor [49], the KASPA [86, 87], the SVETA [85, 88], or the AudioGuider [73].

6.3. Interviews

Although we have already exposed the interviews performed during the design state of this research study in the corresponding chapter, we will now review and summarize the main results and guidelines extracted from them. We will focus exclusively over the aspects and advices regarding the sonification process.

In this line, the most important aspects underlined by the interviewed are, as shown in the corresponding chapter, the following ones:

- Two clearly different timbres. Maybe some simple combinations of a purr with a flute (for example), completing the code later.
- Buzzing is easily perceived.
- Frequency has much information, but we should pay attention to timbres.
- Combination of several sounds/frequencies may bring a lot of information.
- Musical notes might not be useful.
- The sounds and melodies of the mobile phones are good clues, because of their variety, as well as musical instruments.
- A figure should have assigned an involving sound.
- Soft sounds are better, such as human voices, natural sounds, water sounds…
- The set of sounds should not bother. It is important not to use unpleasant or invasive sounds. They should be uncorrelated with the external sources.
- Orchestra instruments are incorporated because of our culture.
- Varying the working cycle of a square wave we can generate many harmonics.
- Unpleasant noise and/or vibration alarms to advice of imminent dangers.
- Boolean alarm to avoid crashes.

Following the results obtained in the Interviews presented and analyzed in chapter 3, some contradictions were found during such interviews. Musical notes were proposed and rejected by different experts. However, the use of culturally or naturally assumed sound sources is seen as the better way to help understanding the acoustic information. Likewise, the frequency is perceived as an important source of information (and other frequency related parameters like the timbre).

Vibrations, alarms or unpleasant sounds were also proposed for dangerous situations. On the other hand, soft sounds should be used for the non-dangerous situations, in order to avoid saturations of the channel or the attention of the user.

We will reproduce the detailed comments and remarks given by the interviewed, already shown in the Interviews chapter.

Distance:

- From further to closer objects, the sound might change (with intermittent beeps, for example).
- Further than 3 meters the information is irrelevant.
- The system might describe the bigger obstacle, or the closer, or the most dangerous ones.
- The depth might be codified in the volume, even if it gives an idea of the drawing and not of the real distance.
- It might add timbres' characteristics. The range should be gradable.

Horizontal axis:

- This axis is related to spatial perception. It is important to use the stereo effects.
- The wide of an object is hard to be imagined. It could be codified with two harmonic sounds.
- The wide of an object could be described with two sounds for the sides of the object.
- Spatiality based on binaural sounds.
- There is a dependency: frequency (f) < 1 KHz, the delay is the most relevant. If f > 1 KHz, the intensity is prevalent. With these aspects, sources are placed in the horizontal axis.

Vertical axis:

- Frequency for the vertical axis.
- Height: children frequencies are higher, adults' and fat people's frequencies are lower.
- The low sounds might be in the bottom, the high at the top of the image.
- In the vertical axis, the head movements are used to place the sound sources.

Temporal variations:

- It could be tested continuous versus intermittent.
- Timbre is useful. A "vibrato" of a musical note could be related with the movement, linking its amplitude and velocity of the displacement.
- When the image is standstill, it should provoke no sounds (implement a differential detector).
- It could be a beep varying its pattern in relation to the depth.

Echolocation:

- It is proposed to use a reference melody and to use the pixels (the depth) as an echo source in a room. Configure the pixels as sources.
- With delays and intensities, objects can be localizable sources in the space.
- Modifying the natural characteristics with the displacement of acoustic sources.
- Taking into account the transmission of the head.

Limits of the system:

- Less than 17 different sounds in any case. We must distinguish between mobility (less than 6 sounds) and vision (less than 17).
- Maybe, it is easier to start with samples, pre-recorded chords, it might be more pleasant and easer.
- The system must have ON/OFF automatic capability: fading in three dimensions.

We find in these proposals contradictions, as well as in the first summary of these remarks. Especially important are those regarding two crucial decisions:

- Psychoacoustic VS arbitrary or mixed sonification paradigm.
- Height translation into frequency.

In the following section, it will be explained and justified the design decisions regarding all these remarks.

6.4. Proposal

The sonification is one of the crucial parts of the proposed system. Indeed, depending on how well this process is implemented, the final assistive product will be accepted or rejected. The users' perceptions, projections and fears must be taken into account (when it is technically and logically possible) because the final acceptation of the product will depend, exclusively, of this group, and not of the technical wellness or efficiency. However, technical aspects are relevant in order to provide answers and solutions to some expressed demands.

6.4.1. Terms

We saw two main paradigms to represent or translate the visual information into sounds, with their advantages and disadvantages. The psychoacoustic approach had advantages regarding the learning, because it exploited the natural localization of sounds. The arbitrary paradigm allowed an extra of information in some specific parameters. Thus, it had been proposed a mix of them, trying to improve both negative aspects. Because of this reason, and given that users have proposed elements of both paradigms, we will finally choose to implement a mixed option.

6.4.1.1. Constrictions of a Real-time Sonification

We have already defined the mathematical representation of a wave. In these terms, we had to deal with periodic additions of sinusoidal waves, following the Fourier transform.

But when we are involved in a real-time system, we have a new aspect to take into account, among others: Real-time constriction means an structural variation or adaptation of the represented information around 20-25Hz (taking the visual constraint).

This means that every 40ms (for the hardest temporal constraint, 25Hz) a new sample of the computed information must be presented. Following this argument, the *piano transform* must be rejected. Let's remember that the vOICe needed around 1 or 2 seconds to represent the scene, frame rate much lower than the real-time constraints. Likewise, other psychoacoustic implementations, based on repetitive patterns like clicks (the EAV [55], for example), will also be rejected.

Another consequence is that no sound below 25 Hz will be reproduced, since the period is larger than the temporal window allowed. This will be applied to the fundamental component of each sound (the lowest one), since the next harmonics are integer multiples of the fundamental one. However, as it has been shown before, the brain presents some capabilities to estimate the fundamental harmonic even when it is not present in the stimulus [43]. Likewise, we found interesting results regarding the rhythm, by Dowling and Harwood [41]. In the interviews, vibrato effects were also proposed to transmit some important (and dynamic information of the scene. However, the rhythm cannot be implemented directly in static image sonification because of the real-time constraints (sonification of one frame), it could be taken into account to represent dynamic changes of the scene (multiple frame or video processing). Anyhow, it will be discussed later in this section.

Kaper and Wiebel [89] describe the Diass, a digital instrument for additive sound synthesis. This application deals with the following parameters of the sound, which will serve as guideline of the possibilities we have to design the sonification code:

Level	Description	Control parameter (S: stationary, D: dynamic)
Partial	Carrier (sine) wave	S: Starting time, duration, phase, D: Amplitude, frequency
	AM (tremolo) wave	S: Wave type, phase, D: Amplitude, frequency
	FM (vibrato) wave	S: Wave type, phase, D: Amplitude, frequency
	Amplitude transients	S: Max size, D: Shape
	Amplitude transient rate	S: Max rate, D: Rate shape
	Frequency transients	S: Max size, D: Shape
	Frequency transient rate	S: Max rate, D: Rate shape
Sound	Timbre	D: Partial-to-sound relation
	Loudness	S: Max size, D: Shape
	Glissando	S: Max size, D: Shape
	Crescendo/Decrescendo	S: Max size, D: Shape
	Localization	D: Panning
	Reverberation	S: Duration, decay rate, mix
	Hall	S: Hall size, reflection coefficient

Table 6.6. Available parameters. We have colored in green those compatible with the real-time constriction. In orange those compatible with a dynamic sonification.

In this table they are present both psychoacoustic and arbitrary characteristics (regarding sonification) of the sound.

Likewise, not every dynamic parameter of the sound will be implemented. Mainly, because they have been rejected by some interviewed (reverberations), some others because they are related with a real-time characteristic and, hence, it will be the scene who creates or not this effect (such as crescendo/decrescendo effects, amplitude modulations, etc.). Other dynamic parameters will be discussed later, such as the vibrato.

Finally, every real-time parameter is important in our approach and all of them will be taken into account.

6.4.1.2. Mapping

Because of the previous discussion and constraints, we will follow the *point mapping* paradigm for the single images sonification:

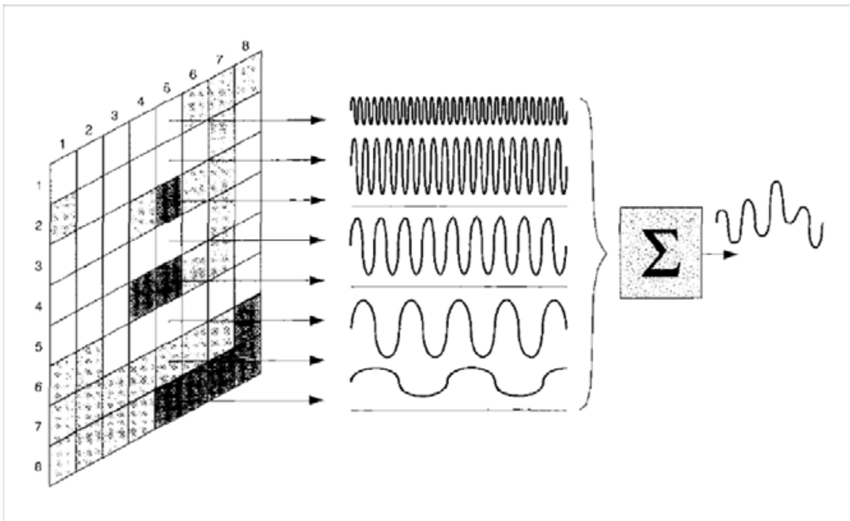


Fig. 6.42. Point mapping paradigm [1].

In this paradigm there was a freedom degree to be set, the meaning of the intensity. As we have already discussed before, the most relevant information for mobility of the blinds, is the distance and position of obstacles. Therefore, the bright, in our sonification input, will represent the distance of every pixel to the user. Moreover, we know we are losing, as remarked Capp [1], plane information such as text. However, this information is not so critical and, in any case, it could be retrieved by means of an OCR routine.

Likewise, it is assumed, following the description of this mapping, that binaurality is used to represent the horizontal position of each processed pixel.

6.4.1.3. Sounds

The simplest sound we can deal with is the sinusoid. This waveform, as explained before, presents only 3 parameters: amplitude, frequency and phase. These simple waves could be used to represent different aspects of the visual world, but, as seen in the interviews, they may be awkward for the user. Thus, some other options appear such as complex real or synthetic sounds.

The discussion about which sounds could be used has been addressed in the available scientific literature. Brown *et al.*, for example, states that musical sounds (instead of pure sine waves tones) are easier to be perceived than the other ones [90].

These musical sounds are much richer in spectral components and, hence, allow us to start talking about timbre. This last parameter helps to segregate different information in a sound stream [91].

Moreover, as stated in [72]:

> "Pure musical tones were deemed more preferable to continuous frequency changes, and synthesized musical instruments preferred over artificial sounds, despite the latter carrying more information through special timbre manipulation".

The only drawback of this option (i.e. musical sounds) is the increase of computational load (if they are produced on-the-fly) or used memory (if they are previously stored).

However, the first two reasons are strong enough to choose this last option as basic sounds, paying the corresponding price depending of the final implementation choice.

Finally, we have to remark that, regarding the real-time constrictions, only stationary sounds can be implemented; this constriction force us to reject, for example, a piano sound, but let us to accept strings, wood or metal wind instruments, or even human choral voices.

6.4.2. Cognitive Aspects

The reception of the sound in the brain, as it was explained previously, is highly dependent of the brain itself, by means of learning processes (training, socialization and other cultural aspects). Castro Toledo defines "Auditory icons as sound pieces, which must be either learned or intuitively understood" [75]. In this research line, Walker [92] defines "earcons" as the acoustic equivalents of icons. These earcons are complex sounds which can be linked together reducing the learning time.

The cultural relation of the understanding is, hence, direct. At this point, we have to wonder how to implement the sonification, taking advantage of culturally available acoustic patterns, regarding both individual sounds themselves (for example, with instruments), as well as the union of different sounds (in an orchestra, chords, etc.).

Likewise, there have been found some relations between cognitive capabilities and acoustic performances. For example, Payne [93] found that pitch discrimination is related with the working memory. However, even if the common sense says that musicians should have better performances when working with auditory displays, Watson and Kidd [94] affirm that this knowledge is not relevant (or a bit relevant in the case of worst musicians and worst non-musicians) when evaluating the auditory perceptual skills.

There is another issue to which we have to pay attention. Persons are unique and, thus, different in several aspects. Those matters with which we have to deal in this work are the cognitive capabilities. As stated in the interviews, the "design for all" paradigm means that everyone should be able to use and join every new device or service, regardless her/his specific abilities. Of course, the way of such use is determined by the design and the ability of the user.

In our proposal, this criterion will be followed implementing a set of progressive profiles, which should give solutions to different needs and levels of detail, abilities, etc. The progressively, hierarchical and additive structure of these levels should help to pass from a level to the next one without needing a new training process, which should be relegated to the first approach to the device. Each new and more detailed level should be complementary, and not supplementary, of the previous ones, providing new information but using the previous one in the same manner as done before.

We have to retrieve an important remark of the experts interviewed: there are relevant and irrelevant information in the environment. Moreover, even if the user, previously trained, is able to distinguish these two sets of information, the possibility of saturating the channel and the attention capability increases if this effort is requested to be done by the user. Thus, the system should be able to discern, in some general cases at least, which information has to be shown and which not, saving important efforts and, hence, easing the understanding of the relevant information; making it clearer.

6.4.3. Relevant and Irrelevant Information

As every other information-based activity, walking and avoiding obstacles is a matter of attention. Among all the things that surround us, we only pay attention to some of them, specifically, the most important ones. This election is done by means of many years of learning process, and some falls, also. But the proposed system doesn't need to understand, as explained already, how is this environment. However, it can discriminate some information which is completely useless regardless the specific situation, as stated in the interviews. Further than 3 meters the information is irrelevant, we found.

Thus, some pixels will be set to black (in an adjustable way by the user), as shown in the next figures.

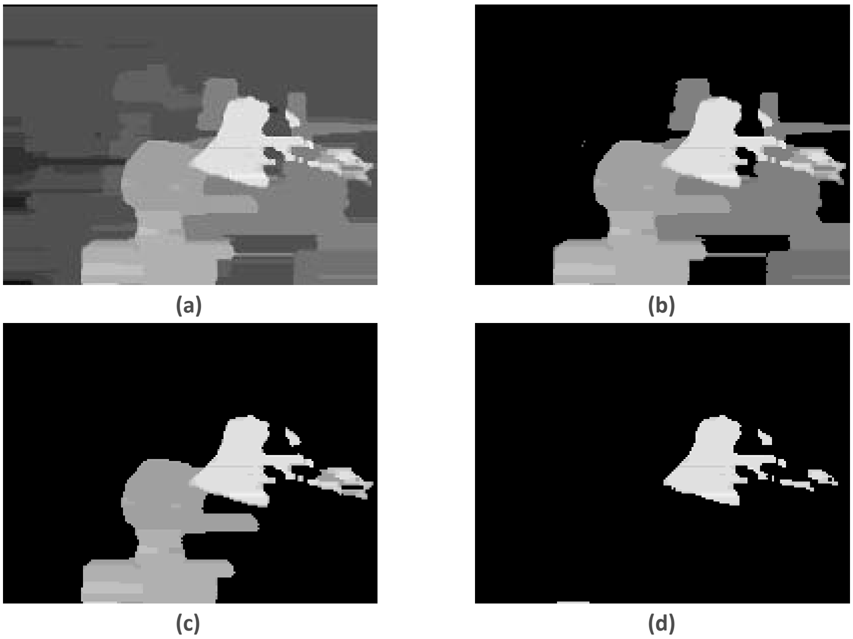


Fig. 6.43. (a) Original 2.5D image. (b-d) A set of truncated images with the irrelevant information erased. (b) Pixels under 100 are erased. (c) Pixels under 150, and (d) under 200 are erased in these two last images.

The threshold which allows pixels representation will be flexible and adjustable by each user and situation.

6.4.4. Sonification Code

All the guidelines given till this point will serve to propose now a sonification code, i.e. a function that relates every relevant characteristic of the image with an identifiable characteristic of the acoustic universe.

In this section we want to be present in the implemented auditory display: horizontal, vertical and depth, for static sonification, and some dynamic representations for moving element in the scene.

The main problem addressed in such definition processes is the choice of the parameters to represent each cue of the 3D-visual world. And, as much as possible, they should follow psychoacoustic properties of the hearing system to simplify the understanding of the code.

6.4.4.1. Horizontal

We saw in the first section of this chapter how the hearing system discriminates the azimuth of the sound source by means of the ILD and the ITD. We saw as well that almost every assistive product proposed used this parameter to distinguish the horizontal position of the sound. Doing so, the user does not have to perform any training regarding this parameter and, hence, the final code is simplified. It will be implemented in this project as stated in the description of the point transform.

However, we have to pay attention to some ambiguities that may arise in this strategy.

In the figure 6.44, the area of the central block (in the left figure) is the double of those areas of the lateral blocks (in the right one). The ILD is exactly the same in both cases, while the scene can be, indeed, the opposite (an obstacle in the middle, two smaller obstacles at both sides). Likewise, only the ITD does not allow perceiving the difference.

We need, then, some other parameter to discriminate between these two canonical situations and, hence, any other similar ones in the real world. A temporal tremolo can represent objects placed in the sides of the image. The depth of the tremolo (the index of the amplitude modulation) will be linked to the horizontal displacement from the central column.

With this decision, the ambiguity proposed in the figure 6.44 disappears: in the first case, the sonification algorithm produces a static sound; in the second one, a tremolo modulates the same sound. Click on the figures to hear an acoustic representation of this parameter.

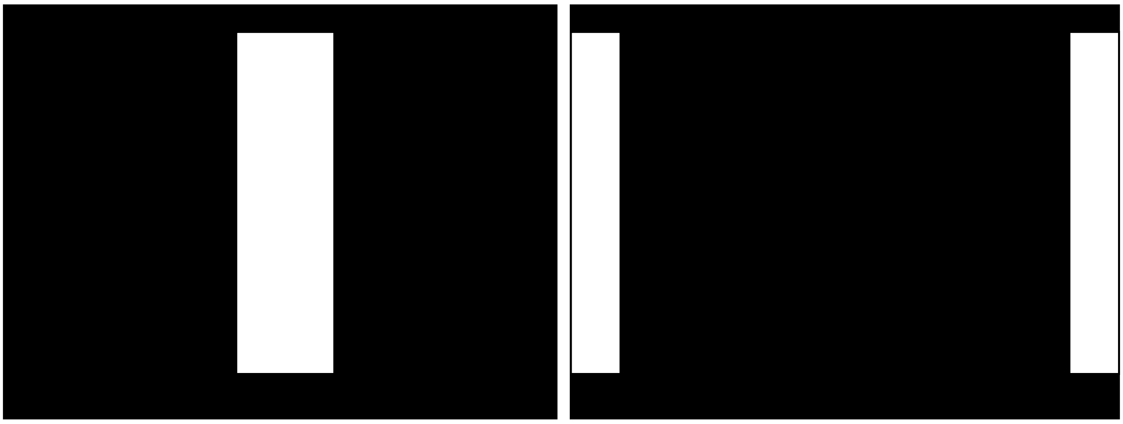


Fig. 6.44. Two situations which are ambiguous if only the ILD and ITD parameters are taken into account.

6.4.4.2. Vertical

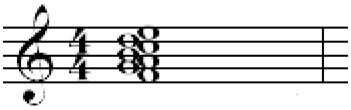
This second parameter is more difficult to be set. It uses to be the one where psychoacoustic devices differ from the mixed ones and the one which is more variable in this last group.

We should take into account the advice of Meijer, stating that the frequency is naturally associated with the high, in a positive way (i.e. the lower is the object, the lower is its frequency, and vice versa) [95].

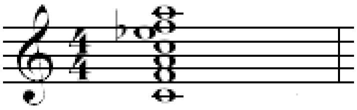
We found exactly the same remark in the interviews. However, we also found the opposite possibility (higher frequencies for those objects lying at the bottom of the image, and vice versa). In our review, we did not find any assistive product or sonification proposal that establishes this last relation between height and frequency, so we will follow the original association.

The point transform (or, in this case, also the piano transform) allows to divide the vertical coordinate in a discrete set of levels. This is mandatory, for instance, because the 2.5D image to be processed is, already, discrete in both dimensions (it is represented by pixels). There are more important advantages of this subsampling. Two tones close enough produce, as shown in figure 6.17, an unexpected new low frequency tone with a pulsation of the difference of the original tones. Thus, if every new tone must be perceived as different of the other ones, we should respect the so-called critical band $\Delta\omega_c$ (figure 6.20). If the separation of each frequency step is big enough (i.e. respecting the critical band), theoretically we will be able to discriminate the sound and, thus, the vertical position, regardless how many tones are present. However, we find a final problem with this option: as shown at the beginning, the addition of several uncorrelated tones produce noise. Perfectly we could propose to use narrowband noise to represent how high a detected object is, but from the interviews we can extract some other conclusions. It was expressed in several cases the problem of hearing disagreeable or buzzing sounds all the time the device is in use. Thus, the noise option should be rejected.

There is another possibility: to use musical chords and octaves. We found something similar in an already presented project, the SVETA [81]. Moreover, the researchers of this project affirm that the central octave (i.e. that whose range is 440 to 880 Hz) is more pleasant. Likewise, the separation between adjacent tones is performed, as shown in eq. 6.14 and the following explanation, with musical tones. More specifically, it is done using a major chord (two tones between the base and its third note, and 1.5 tones between the third and the fifth note). With these notes, they produce eight levels including some major chords, subsampling the 32 rows image by a factor of 4. The advantage of this implementation (which is, however, not clear in some aspects in the published works) is that the addition of harmonic notes (the fundamental plus its third, fifth, octave and so on) does not produce the sensation of hearing noise, but a harmonic union of sounds, i.e. musical chords:



(a)



(b)

Fig. 6.44. (a) Uncorrelated notes and (b) a major chord. Both of them present the same number of notes. Ctrl+Click in the chords to listen them.

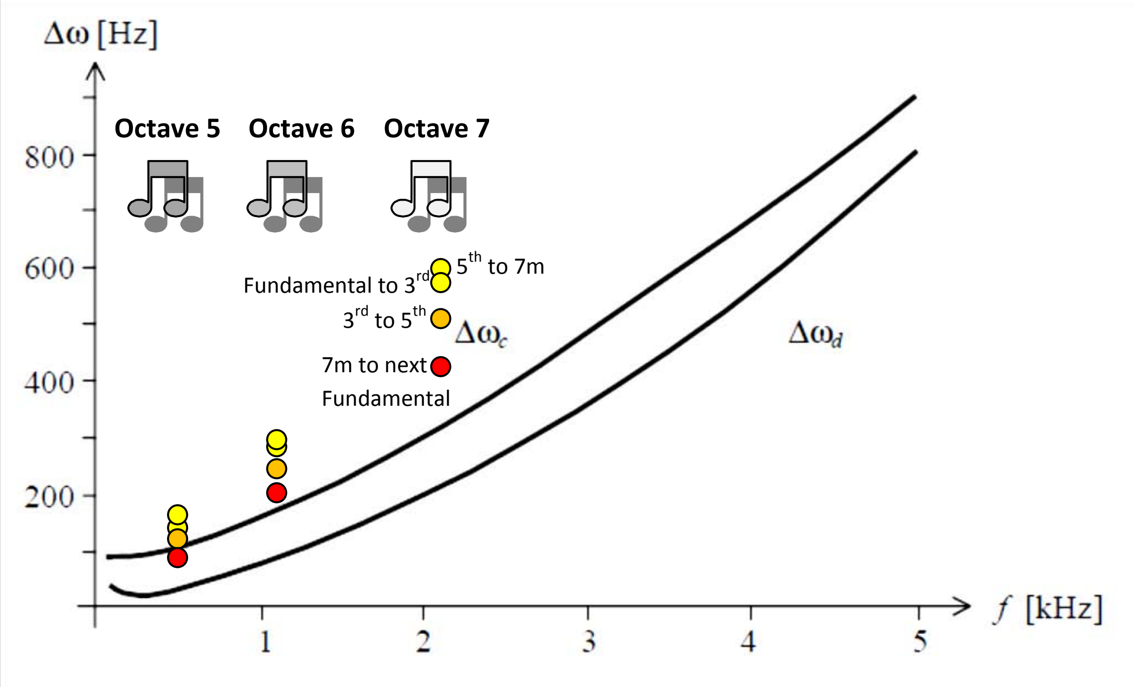


Fig. 6.45. The intervals of fig. 6.20 represented over the critical band curves for the three highest octaves. In yellow, the two bigger differences, i.e. 5th to 7m and the fundamental to the 3rd, in orange the 3rd to the 5th and, finally, in red, the smallest interval, 7m to the next octave fundamental note. These points are placed representing the intervals of the CMaj7m chord. Ctrl+Click on the musical symbols to hear each chord.

Even if some intervals in lower octaves may lie under the theoretical limit of the critical band, which must be respected to perceive completely separated tones, but over the frequency discrimination threshold, or $\Delta\omega_d$, we can still perceive different tones.

However, given that the BC imposes restrictions above 10 KHz, the 5th octave will be the highest one used in this study.

Note	Octave 2 [Hz]	Octave 2 Δω [Hz]	Octave 3 [Hz]	Octave 3 Δω [Hz]	Octave 4 [Hz]	Octave 4 Δω [Hz]	Octave 5 [Hz]	Octave 5 Δω [Hz]	Octave 6 [Hz]	Octave 6 Δω [Hz]	Octave 7 [Hz]	Octave 7 Δω [Hz]
C	65.41	3.89	130.81	7.78	261.63	15.55	523.25	31.11	1046.5	62.23	2093	124.46
C#	69.3	4.12	138.59	8.24	277.18	16.48	554.36	32.97	1108.73	65.93	2217.46	131.86
D	73.42	4.36	146.83	8.73	293.66	17.47	587.33	34.92	1174.66	69.85	2349.32	139.7
D#	77.78	4.63	155.56	9.25	311.13	18.5	622.25	37.01	1244.51	74	2489.02	148
E	82.41	4.9	164.81	9.8	329.63	19.6	659.26	39.19	1318.51	78.4	2637.02	156.81
F	87.31	5.19	174.61	10.39	349.23	20.76	698.45	41.54	1396.91	83.07	2793.83	166.13
F#	92.5	5.5	185	11	369.99	22.01	739.99	44	1479.98	88	2959.96	176
G	98	5.83	196	11.65	392	23.3	783.99	46.62	1567.98	93.24	3135.96	186.48
G#	103.83	6.17	207.65	12.35	415.3	24.7	830.61	49.39	1661.22	98.78	3322.44	197.56
A	110	6.54	220	13.08	440	26.16	880	52.33	1760	104.66	3520	209.31
A#	116.54	6.93	233.08	13.86	466.16	27.72	932.33	55.44	1864.66	110.87	3729.31	221.76
B	123.47	7.35	246.94	14.68	493.88	29.38	987.77	58.73	1975.53	117.47	3951.07	234.93

Table 6.7. Intervals among every note of the well-temperate scale.

Note	Octave 2 [Hz]	Octave 2 Δω [Hz]	Octave 3 [Hz]	Octave 3 Δω [Hz]	Octave 4 [Hz]	Octave 4 Δω [Hz]	Octave 5 [Hz]	Octave 5 Δω [Hz]	Octave 6 [Hz]	Octave 6 Δω [Hz]	Octave 7 [Hz]	Octave 7 Δω [Hz]
C(i)	65.41		130.81		261.63		523.25		1046.5		2093	
E	82.41	17	164.81	34	329.63	68	659.26	136.01	1318.51	272.01	2637.02	544.02
G	98	15.59	196	31.19	392	62.37	783.99	124.73	1567.98	249.47	3135.96	498.94
A#	116.54	18.54	233.08	37.08	466.16	74.16	932.33	148.34	1864.66	296.68	3729.31	593.35
C(i+1)	130.81	14.27	261.63	28.55	523.25	57.09	1046.5	114.17	2093	228.34	4186	456.69

Table 6.8. Intervals among the notes of the CMaj7m chord

6.4.4.3. Depth

The depth is one of the most important cues for a mobility activity. Thus, we will leave apart the advice of Capp [1], talking about the loss of flat information.

Given that the basic images over which we work at this point represent the depth, the brightness is then logarithmically related to the distance. Moreover, in the point transform, this brightness was proportionally related with the volume. As stated in the SVETA project, "The volume of the sound increases with the increase in disparity" [81].

However, the volume will not act as an absolute distance measurement. As every real-world application, the system must be adaptable to different situations, and inside a building the volume needed to hear the sonification is much lower than walking in the street. Thus, regardless the system implements an automatic gain control, or this adaptation is done manually, the volume does not represent in an absolute manner the distance any more. Thus, an additional parameter is needed to allow safe distance perceptions.

As we saw in the subsection dedicated to sounds, Cusack and Roberts [91] reported the utility of the timbre to discriminate information embedded in an acoustic stream. We will then codify the distance (brightness) of a represented pixel by means of the timbre. We propose, in this line, that the far pixels will be represented by woodwind sounds, such as a flute or oboes, and this parameter becomes sharper (with a soft evolution until the trumpet, for example, in the nearest ones) with the increase of the brightness.

A test button, to hear the maximum amplitude in each state of the system can also be implemented.

6.4.5. Software

The techniques to converting a grey scale image to a set of sounds can be gathered in, at least, three different ways:

1. Addition of on-the-fly generated sounds.
2. Multi-band filtering of white noise.
3. Sum of weighted pre-generated sounds.

All these possibilities have advantages and disadvantages:

1. In the first case, the consumption of memory is very low. Only some parameters to control the way in which sounds must be generated is needed. On the other hand, generating N×M sounds (being N×M the dimensions of the image; however, it can be resized for this purpose) is expensive computationally. There is a sum of every single wave. "With a sampling rate of 44.100 samples per second, the time available per sample is only 20 microseconds—too short for real-time synthesis of reasonably complex sounds. For this reason, most of today's synthesis programs generate a sound file that is then played through a DAC" [89]. We will see there are, nowadays, some possibilities in this field.
2. In the filtering option, memory is, again, quite low, although not as low as the previous case. Generating a white and Gaussian noise is not expensive computationally, but

filtering this sound by N×M bands, which must be sharpen enough, is a very costly process, mucho more expensive than the previous one.

3. Improving the performance of the first case, but increasing the amount of memory used, we can find a solution to this problem by pre-defining a N×ML 3D matrix (L is the length of the so-called wavetable) were it is stored, before the use of the real-time system, the whole set of waves that can be used by this system. The final addition of the generated waves is, again, necessary.

We can summarize the performances of each option in the following table.

Case	Computational cost	Memory requirements
On-the-fly generated sounds	↑↑	↓
Multiband filtering	↑↑↑	↑
Pre-generated sounds	↓	↑↑

Table 6.9. Comparison of the three options in terms of computational load and memory requirements.

The second option is not interesting, since the computational cost is very high, much more than any other option. Another disadvantage of this second option is that it works filtering wide band noise. But there are many research works that, as already seen, demonstrate the advantage of structures sounds, even musical scales. Moreover, in the interview, we found a preference to culturally learned sounds, even natural sounds like human voices. All these reasons force us to reject this option and to focus the efforts in someone of the other two.

These other two options are based on the so called sequencer. A sequencer generates or plays sounds of specific characteristics when it is required. The information that should determine which sound, how and when it is reproduced will be each represented pixel of the image.

There is a standardized sequencer protocol, based on synthetic generated sounds (but applicable to pre-generated sounds): the Musical Interface Digital Interface or MIDI [96].

6.4.5.1. MIDI

MIDI is a well established standard for communications among digital musical instruments. It is, thus, music oriented.

MIDI is proposed in 1983 by a holding of enterprises of the musical industry as a standard to make compatible all the sequencers [97]. This first protocol, the so-called MIDI 1.0, specified the structural aspects (physical and logically) of the MIDI communication, but with some ambiguities which forced, in 1990, to define the General MIDI standard, with the following minimum capabilities [97]:

- Multitimbric (several and simultaneous reproduction of different instruments) capability of 16 channels.
- Minimum polyphonic (simultaneous notes, independently of the number of instruments) capability of 24 notes.
- Standard map of 128 instruments.
- Drum box in the channel 10 with, minimum 59 percussion sounds.

- Pan and loudness.

We can find limitations, in applications to sonification: “using only MIDI notes 35-100. Even in that case, the designer has only 65 display points to use to represent whatever data they may have” [11]. However, this does not take into account the rest or parameters we can play with.

Moreover, this standard has been extended with the GS Extension and, finally, in 1999, the General MIDI Level 2, or GM2 [96]. The main new features of this protocol are summarized in the following list [96]:

- Number of Notes - minimum 32 simultaneous notes
- Simultaneous Percussion Kits - up to 2 (Channels 10/11)
- Up to 16384 variation banks are allowed, each containing a version of the 128 Melodic Sounds (the exact use of these banks is up to the individual manufacturer.)
- 9 GS Drum kits are included
- Additional Control Change messages:
    - Filter Resonance (Timbre/Harmonic Intensity) (cc#71)
    - Release Time (cc#72)
    - Attack time (cc#73)
    - Brightness/Cutoff Frequency (cc#74)
    - Decay Time (cc#75)
    - Vibrato Rate (cc#76)
    - Vibrato Depth (cc#77)
    - Vibrato Delay (cc#78)
- Registered Parameter Numbers (RPNs)
    - Modulation Depth Range (Vibrato Depth Range)
- Universal SysEx messages
    - Master Volume, Fine Tuning, Coarse Tuning
- Reverb Type, Time
- Chorus Type, Mod Rate, Mod Depth, Feedback, Send to Reverb
- Controller Destination Setting
- Scale/Octave Tuning Adjust
- Key-Based Instrument Controllers
- GM2 System On SysEx message

The MIDI standard defines both physical and logical levels for the communication compatibility. Likewise, these two levels are independent, and since we will transmit the sonification information through a serial port (see next chapter), we will only focus on the logical level.

This level is defined by a set of messages, which present the following structure:

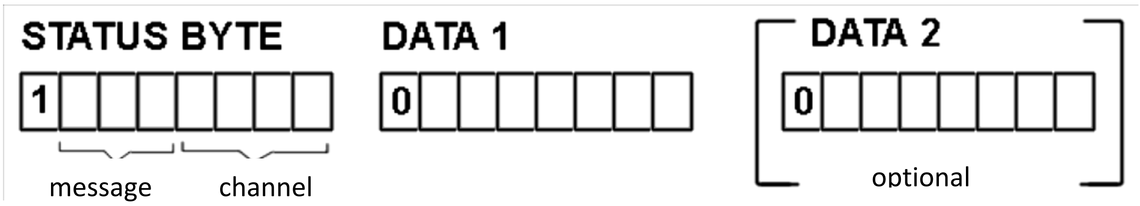


Fig. 6.46. MIDI message structure [97].

We have, thus, 8 messages, and the parameters are ranged from 0 to 127.

The complete set of messages of the specification General MIDI (1.0 and 2) is shown in the tables a1.1, a1.2, a1.3 and a1.4 of the Annex I.

As it can be appreciated in these tables and figure 6.46, MIDI has 16 channels to be free and independently configured. This is a constraint of the number of instruments and sound parameters, which would be able to be reproduced:

- Each channel can only be configured once for each simultaneous note.
- Hence, the panoramic parameter must be set for each channel at the beginning.
- The same constraint is applied to the vibrato effect and to the instrument.
- Nowadays, MIDI hardware is able of reproducing up to 128 simultaneous voices (not instruments, whose limit is 16), being 64 the average amount for cost-effective devices.

The limitations in the MIDI sequencers force us to op between two possibilities:

- Process the image simplifying it to N×8, or
- Process the image simplifying it to N×16.

The reason of this limitation is the number of independent channels. Having only 16 channels, we will have to divide the image in 16 or 8 columns. On the one hand, if we use 16 channels with different panoramic values, in each column will be assigned to a channel and, thus, will have a specific value of panoramic. This option presents the drawback of having only one instrument in each column. This occurs independently of the number of rows, since the same instrument can play different notes without "consuming" more than one channel.

The vision cone is approximately $90^{o}\times90^{o}$. In our case, the visual representation has a static positioning step in the azimuth of

$$\varepsilon = \frac{90^{0}}{16*2} = 2.81^{0} \tag{6.15}$$

On the other hand, if we assign a pair of channels to each column, in the case of 8×8 or 16×8, we will have two channels for each panoramic (and vibrato) value, but two instruments per column. The same step in this case is:

$$\varepsilon = \frac{90^{0}}{8*2} = 5.625^{0} \tag{6.16}$$

We will reproduce sounds with panoramic values between "0" and "FF" in MIDI code, which represent an audition field of 180^{o}. The main reason of such wide amplitude is the mutual interference that the bone transmission produces in both ears. Thus, separating the sources as much as possible will produce a clearer stereo effect. There is no reason to choose the first option, and the image will be processed in N×8 pixels format. Moreover, each column will be represented by two channels, reproducing each one of them a different instrument, if needed. The step is between 18 units, thus, 10^{o}. We will have, then, an error around 5^{o} in the position estimation, lying around the theoretical limit of the human hearing system for the static representation of acoustic sources, in table 6.3.

A final remark regarding the channels must be done: channel 9 is reserved for a drum set, and it must be avoided.

Regarding the value of N (rows), it can be defined dynamically depending of the cognitive necessities or limitations. We will discuss this matter in the following section.

Table A.I.4 resumes the instruments defined in GM1. From these instruments, we will select a set of 5 to represent the depth (given that the depth is, as well, represented by the volume) ordered from softer to harder timbres. These instruments are exposed in the table 6.38 and the sound corresponds to the central C.

Instrument #	Instrument Name	Sound (Ctrl+Clik)
0x36	Synth. Voice	♫
0x49	Flute	♫
0x44	Oboe	♫
0x39	Trombone	♫
0x3B	Muted Trumpet	♫

Table 6.10. Selected instruments.

From these tables, we will use the following messages, regarding the proposed sonification:

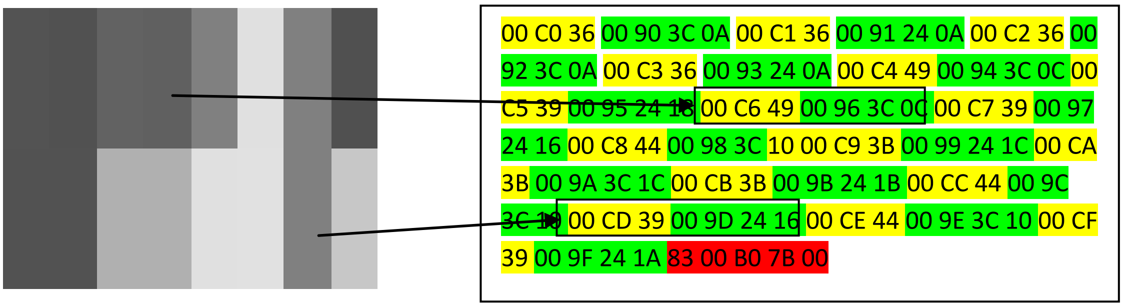
- "Note On": This message activates a note, with the following code: *{0x90+pan, 0xnn, 0xvv}*, where the "pan" parameter is the panoramic value, which is linked to a channel (thus, value ranged from 0 to F), "nn" is the note, regarding the table A.I.3, and "vv" the volume, a fraction of the depth to avoid saturation when many notes are simultaneously being reproduced.
- "Vibrato set": This message is sent only once, at the initialization process, to assign to each channel a vibrato value: *{0xBn, 0x4C, 0xvv}* and *{0xBn, 0x4D, 0xvv}*, where "n" represents the channel (from 0 to F) and "vv" the vibrato rate in the first sequence and the depth in the second one.
- "Pan set": The "pan set" message allows assigning a panoramic value to each channel: *{0xBn, 0x0A, 0xvv}*, where "n" is again the channel and "vv" the panoramic localization ("0" represents the left, "FF" the right).
- "Set instrument": This message changes the program and, thus, the instrument played in a channel: *{0xCn, 0xvv}*, where "n" is the channel, and "vv" the value, computed by means of a look-up table (LUT) with the following values: {*0x36, 0x49, 0x44, 0x39, 0x3B, 0x3B}*. This table is accessed with the following index, only computed if the intensity of the corresponding pixel is higher than 42:

$$i = \frac{value}{42} - 1 \tag{6.17}$$

  This index, combined with the LUT, assigns the value of the 255 ranged intensity of a pixel to the corresponding instrument.

- "All notes off": to turn off all the notes, for example when we are looking for examples of 1 second, there is this special message: *{0xB0, 0x7B, 0x00}*

Thus, here we can find an example of the instructions computed for a quite simple image:



```
00 C0 36 00 90 3C 0A 00 C1 36 00 91 24 0A 00 C2 36 00
92 3C 0A 00 C3 36 00 93 24 0A 00 C4 49 00 94 3C 0C 00
C5 39 00 95 24 18 00 C6 49 00 96 3C 0C 00 C7 39 00 97
24 16 00 C8 44 00 98 3C 10 00 C9 3B 00 99 24 1C 00 CA
3B 00 9A 3C 1C 00 CB 3B 00 9B 24 1B 00 CC 44 00 9C
3C 10 00 CD 39 00 9D 24 16 00 CE 44 00 9E 3C 10 00 CF
39 00 9F 24 1A 83 00 B0 7B 00
```

Fig. 6.47. Two vertical levels image and the corresponding MIDI code (in hexadecimal base) of a Standard MIDI File (SMF). The channels configuration is not shown in this code. In yellow, the "program change" (what we called "set instrument") messages, in green the "note on" messages and in red the "all notes off" message. With arrows, the correspondence of some pixels and their MIDI code. Ctrl+Click on the image to hear the generated sound.

6.4.6. Profiles

As we already said, design for all means the capability of adaptation to different cognitive and behavior patterns and skills. This constraint forces us to develop a set of different and progressive profiles. The idea was already shown in the figures 4.5 and 4.6. If every new and more complex profile is just an addition of information regarding the previous one, the learning process will be dramatically reduced.

For each profile, there are some parameters adjustable by the user, such as distance threshold and volume. The basic image which will be used for sonification has 16x8 pixels.

We propose, finally, the following set of additive profiles.

6.4.6.1. Profile 'P0'

The most basic information is the bit, i.e. a Boolean alarm which alerts or not the user depending on the obstacles. This alarm should be implemented both in a vibratory device as well as with an unpleasant sound. This sound will be placed in front of the user and in the central frequency of the system.

The number of pixels above the closeness threshold is an arbitrary choice. We will set this parameter as 2000 pixels over the threshold, to activate the alarm. In the following figures we can see this number is big enough to avoid false alarm effects, but small enough to alert of some small obstacles. The closeness threshold is set, for representation purposes, in 200.

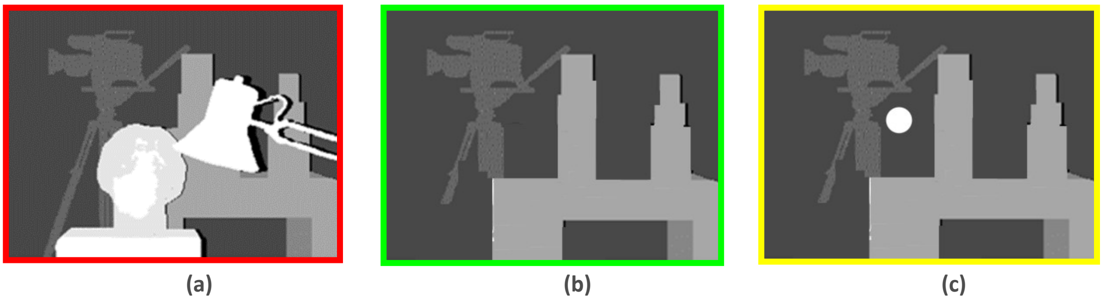


Fig 6.48. Examples of (a) Alarm ON (86739 pixels over the threshold), (b) Alarm OFF (295 pixels over the threshold) and (c) limit of the alarm (2794 pixels over the threshold value) images. Ctrl+Clik on the images to listen the alarm, if it is the case.

This profile is designed for deafblind people, as well as for blind people in very noisy environments. Moreover, people with some cognitive limitations will also be able to use the basic profile.

The vibratory alarm (profile ‘0’) should be able to be added or switched off independently in the rest of levels.

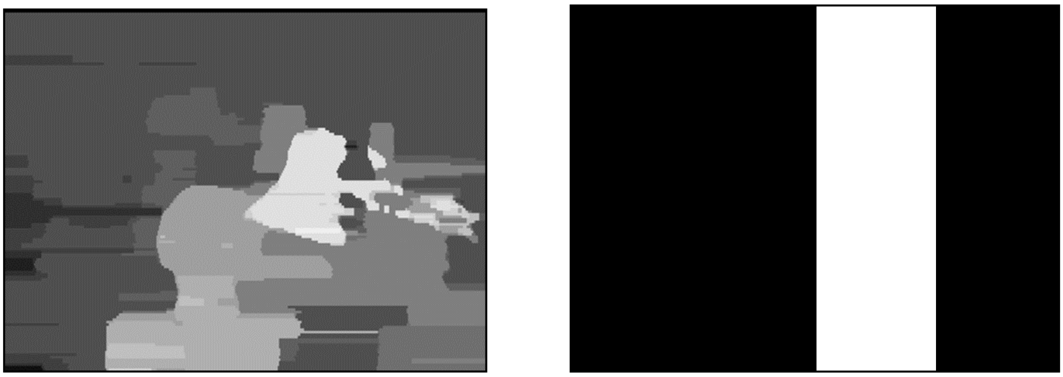
6.4.6.2. Profile ‘P1’

From this level on, we start to talk about sonification based profiles.

The basic image to be sonified will have only Nx8 pixels, as discussed before. For instance, the number of rows will be also 8.

This level implements, again, a Boolean alarm, but it allows to localize it horizontally. This positioning will be done, as explained, by means of the ILD and ITD, as well as with a vibrato effect. The sound will only reproduce the most strident one. Thus, no measure of distance is given.

For this image, if there is a pixel whose value is over the threshold, the corresponding vertical line is activated.



(a)

(b)

Fig. 6.49. (a) Original low resolution image and (b) White vertical lines where obstacles are detected. Ctrl+Click on the (b) image to hear the corresponding sonification.

In this image, the threshold of the pixels value has been set to 200 and the number of columns, as said before, reduced to 8.

6.4.6.3. Profile ‘P2’

This new level reproduces the information given by the previous one, but inserts new volumes and timbres, so it allows a depth perception. The frequency is still placed in the middle of the band.

For the same image, the intensity threshold is reduced to 42, and each distance is represented by different instruments, so more information is available. Finally, in each column, the maximum value determines the value of the whole column. Thus, it sonifies the worst case for each column.

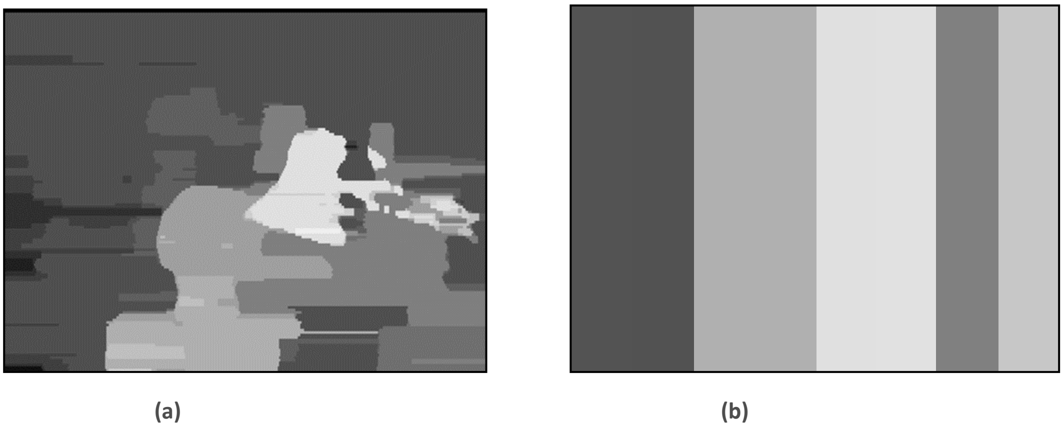


Fig. 6.50. (a) Original low resolution image and (b) Gray vertical lines where obstacles are detected. Click on the (b) image to hear during 1 s the corresponding sonification.

6.4.6.4. *Profile 'P3'*

Profiles '3' to '6' are successive vertical partitions of the image, introducing in each new level, the double of sounds present in the previous one. In this first step, the image is divided into two vertical subimages. Each one of them is represented by a different frequency: C_2 and C_4.

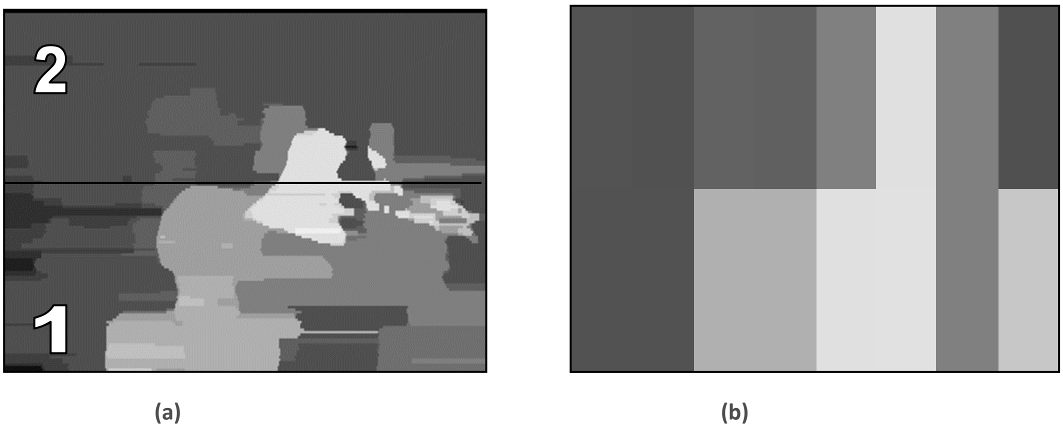


Fig. 6.51. (a) Original low resolution image and (b) vertical lines representation of each semi-image. Click on the (b) image to hear during 1 s the corresponding sonification.

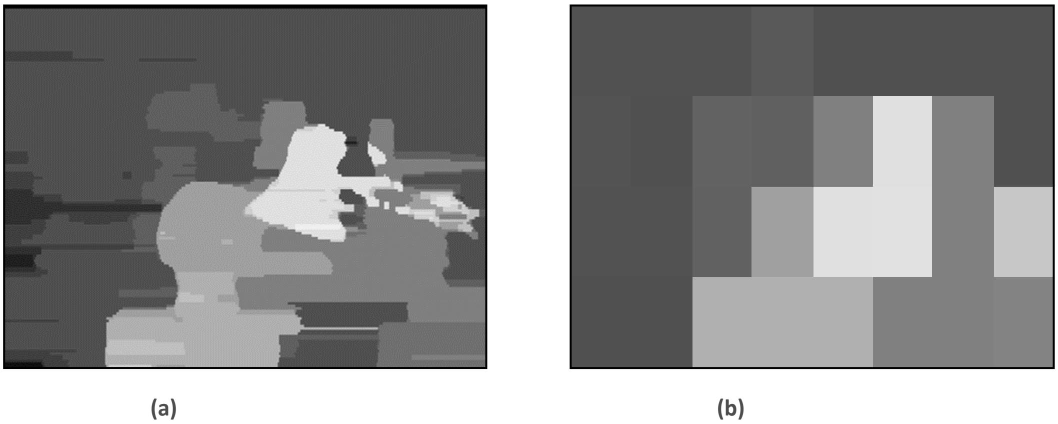
Since we have two levels of height, and two channels per column, in this level, the upper and the lower part of the sonified image (fig. 6.51.b) are synthesized in independent channels and, hence, with their own instrument.

All the other parameters stay stable.

6.4.6.5. *Profile 'P4'*

Every section of the previous profile is divided in two.

Thus, four frequencies are used in this step, all the fundamental notes: C_2, C_3, C_4 and C_5.



(a) (b)

Fig. 6.52. (a) Original low resolution image and (b) vertical lines of the sonified image. Click on the (b) image to hear during 1 s the corresponding sonification.

From this level on, we can observe a limitation in the sonification code. The MIDI protocol does not admit more than one instrument per channel, and we have 16 channels grouped by pairs. Then, the most optimal choice is to order the instruments (gray levels) per column, and apply the following rule:

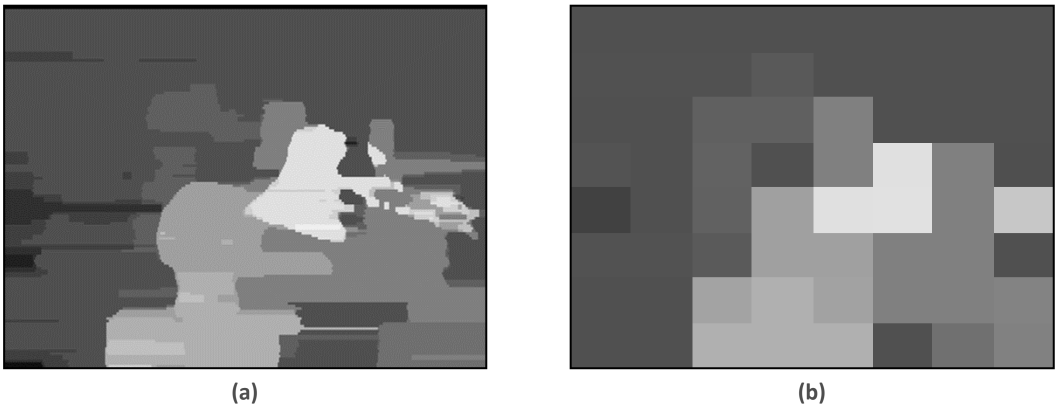
- If there is only one depth value per column (or if all the depth values stay in a range small enough), assign the corresponding instrument to each pixel.
- If there are two different depth range values, apply two instruments, each one to the corresponding pixel.
- If there are more than two different range values, sonify the highest one (the closest pixels) with the corresponding instrument. Then, sonify the other ones with the second higher instrument.

Let's remark that, independently of the instrument used for each depth range, the value of the pixel, as told before, determines the volume of the corresponding note.

This rule will be applied from now on in the rest of profiles.

6.4.6.6. Profile 'P5'

Profile '5' doubles the number of used frequencies. The 5th frequencies G_2, G_3, G_4 and G_5 represent the 4 new vertical intervals.



(a) (b)

Fig. 6.53. (a) Original low resolution image and (b) vertical lines of the sonified image. Click on the (b) image to hear during 1 s the corresponding sonification.

6.4.6.7. Profile ‘P6’

The profile ‘6’ is the most complex one, respecting the limit retrieved from the interviews of less than 17 sounds. Thus, 16 frequencies are used. The 8 new frequencies regarding the previous profile are the 3rd and the 7thm, i.e., E_2, Bb_2, E_3, Bb_3, E_4, Bb_4, E_5 and Bb_5.

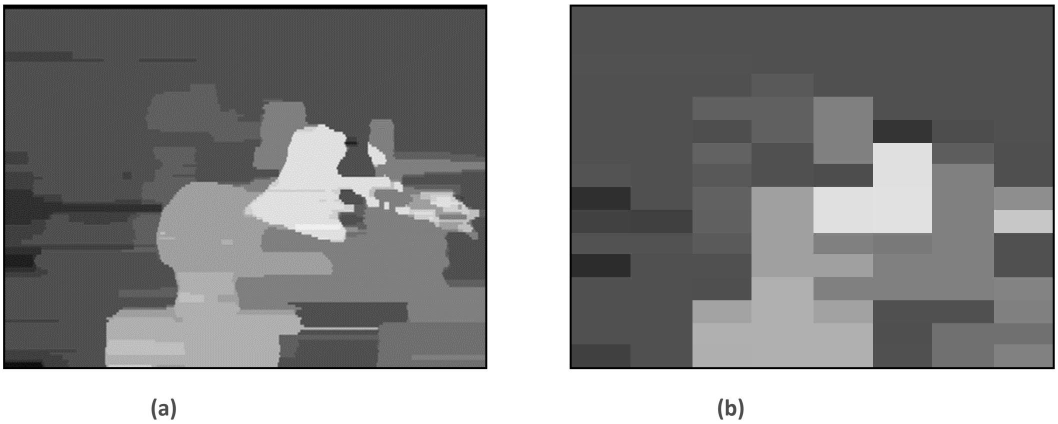


Fig. 6.54. (a) Original low resolution image and (b) vertical lines of the sonified image. Click on the (b) image to hear during 1 s the corresponding sonification.

The problem we have to deal with in this profile is the limitation of simultaneous voices or notes that a MIDI synthesizer can play at a time. Although we can find, as it will be discussed in the following chapter, hardware capable of 128 voices, the complexity and price of this hardware present some inconvenient for our purpose. Thus, we will have to retrieve the most relevant information from a 16×8 pixels image; in fact, the 64 more relevant pixels.

This will be done ordering them with the intensity as criteria, and sonified in this order, as it can be seen in the following video:

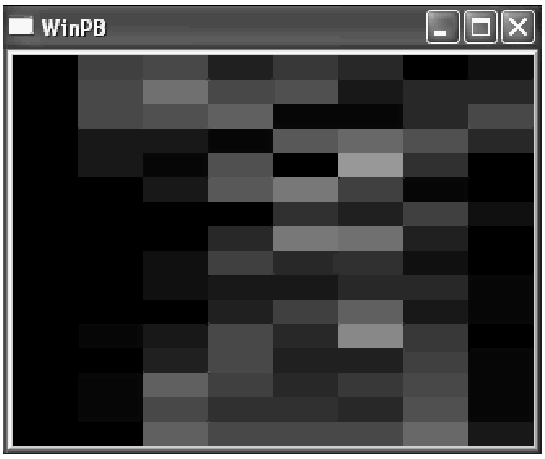


Fig. 6.55. Video of the ordered selection of pixels. Click on the image to watch the video.

6.4.7. Video Constraints

All the precedent images have been computed for static pair of images. However, our proposal has real-time constraints. The main problem we have to face to is being able to transmit all the information needed through the relatively slow MIDI protocol. This limit is set in the standard as 3125 bytes/sec [97].

We take the worst case into account (miscounting the set-up process of pan and tremolo variables):

- 16×8 pixels illuminated (truncated or not to the 64 more brilliant)
- 5/6 bytes per pixel to code a MIDI message
- 25 fps

This case needs 10-19.2 KBps, 3.2-6.1 times higher than the supported bandwidth in the GM specification. Let's notice that each byte is transmitted with a start and end bits, so 10 bits per byte are required.

However, internal communication of the MIDI messages in the computer is not limited to the MIDI specifications. Moreover, new hardware can manage faster communication protocols, what will be discussed in the next chapter.

To simplify the scenario, we will assume the following supposition: Any real-life situation changes softly in the temporal dimension. Other simplifications can be done using the flexibility of the MIDI protocol, as it will be explained in the following section.

6.4.8. PC Version of the Real-time Sonification Protocol

Standard PCs can generate, process and reproduce MIDI messages in real-time environments. Windows is equipped by default with a GM1 player (the Windows MIDI Mapper [98]), with quite limited capabilities. This driver has shown to ignore the aftertouch message, the vibrato and tremolo messages, the brightness messages, among others. Thus, although for file based MIDI processing (as done up to this point), it is useless for real-time processing. Changes of the instruments, note-on messages every 40ms and vibrato emulated implementation by means of changing the pitch of the channel creates an incomprehensible noise, shown in the following video.

Fig. 6.56. First real-time sonification version. Click on the image to watch the video.

Thus, a GM2 capable synthesizer must be used to avoid unnecessary internal MIDI traffic, as well as a better and softer sonification.

The Edirol HyperCanvas HQ GM2 [99] by Roland, is able to process most of these messages. An important drawback of this synthesizer is that it does not respond to the polyphonic aftertouch messages, which allow changing individually the volume of a specific note, without needing to turn it off and again turn it on with the new volume. The solution to this problem will be shown later.

This synthesizer is a Virtual Studio Technology [100] and, hence, needs to be driven by a VST manager. The choice was the V-Stack program [101]. Finally, a pipe must be implemented between the Windows Multimedia library and the V-Stack. This pipe was done with the MIDI Yoke system [102].

However, as said before, the volume change of each note is ignored by the Edirol synthesizer. It is not ignored, on the other hand, the channel volume. Given that the MIDI Yoke implements 8 MIDI pipes, a new mapping of the 16×8 image can be done, using one note per channel and, hence, allowing changing their volume individually:

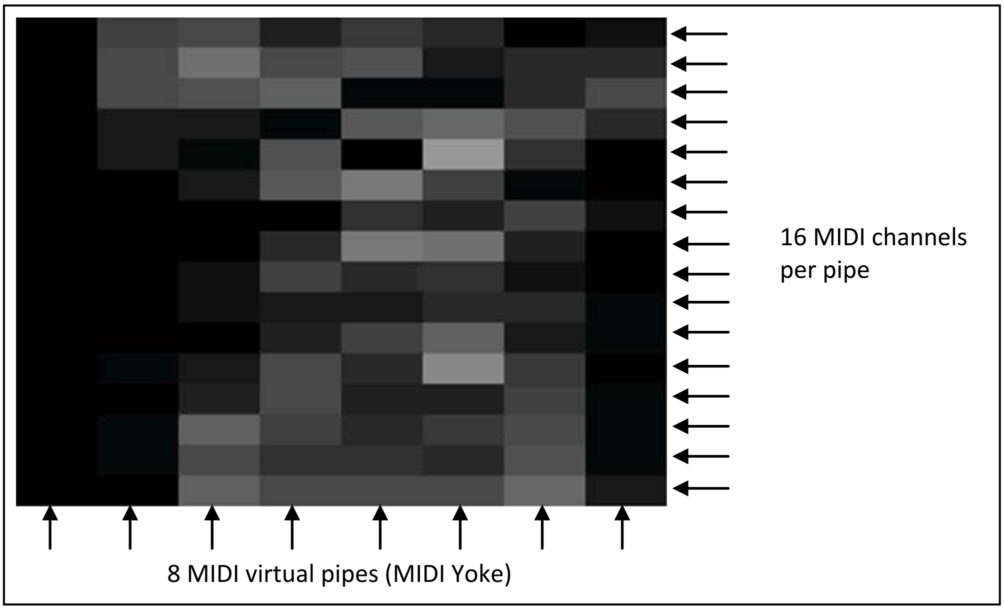


Fig. 6.57. Mapping of the GM2 based sonification.

Another problem to be focused is the impossibility of changing the instrument in real-time without provoking a click (given that the attack time of the new note is not null). We find in the GM2 MIDI standard the brightness control of a sound, {0xBn, 0x4A, 0xvv} where “vv”, the cut off frequency of the low pass filter of the MIDI synthesizer, varies between 25 and 100. The basic instrument is the trombone, with rich high frequency harmonics.

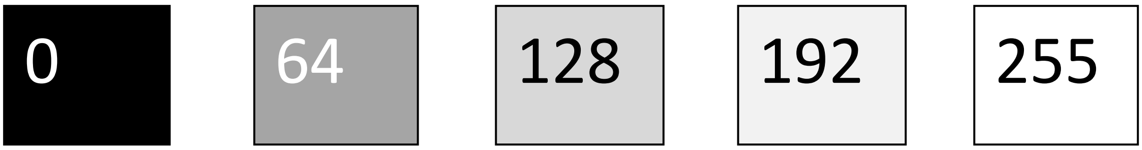


Fig. 6.58. Examples of cut off frequencies for different pixel intensities. Click on the images to hear.

This effect is combined with the volume, varying from 0 to 0x7f (127 in decimal notation).

The final flow chart of the MIDI information inside the computer was set as follows:

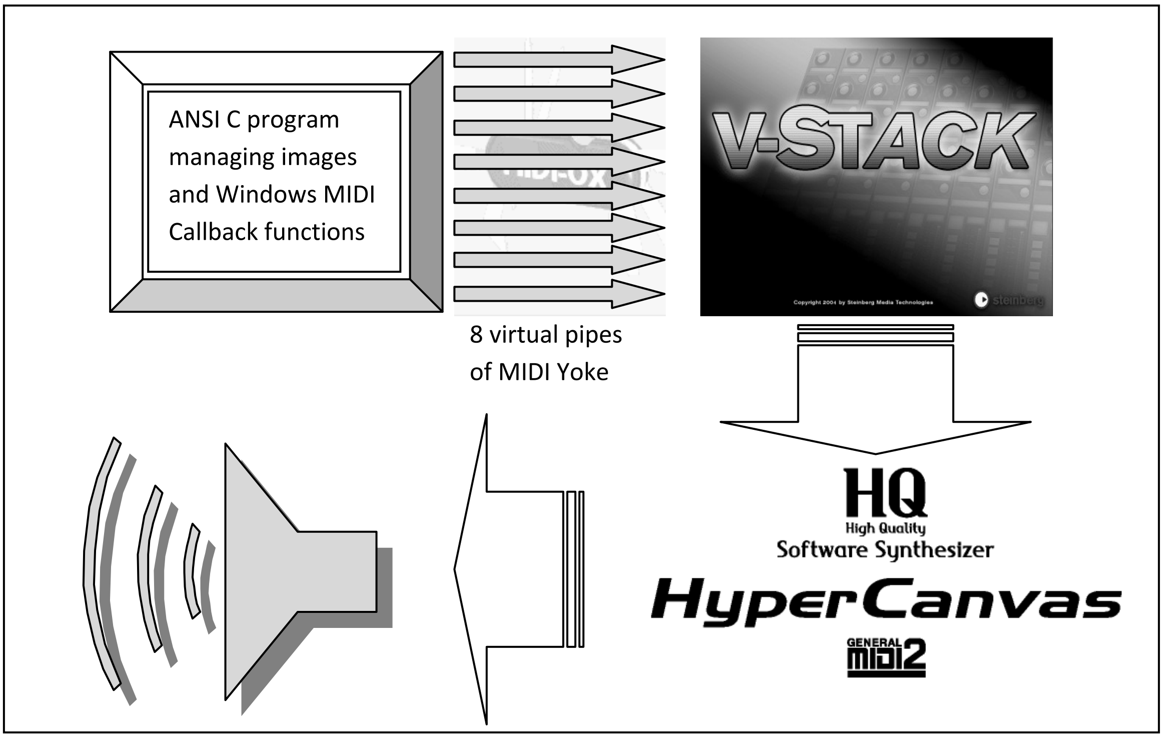

ANSI C program managing images and Windows MIDI Callback functions
8 virtual pipes of MIDI Yoke
V-STACK
HQ
High Quality
Software Synthesizer
HyperCanvas
GENERAL midi2

Fig. 6.59. Flow chart of the GM2 MIDI messages in the PC second version.

The initialization of the MIDI system is done with the following messages, and once per use:

```
//---------MIDI system opening, 8 pipes with 16 channels each-----------------------------------
for( int i = 0; i < 8; i++){
                        (int) midiOutOpen(&lphmo[i], i, 0, 0, 0);
                        for( int j = 0; j < 16; j++)
                                    midiOutShortMsg(lphmo[i],head[i]+j); // Channel panoramization
            }
//--------Instruments initialization-----------------------------------------------------------------------------
for( int i = 0; i < 8; i++)
                        for( int j = 0; j < 16; j++){
                                    midiOutShortMsg(lphmo[i],0x3900+0xc0+j);
                                    midiOutShortMsg(lphmo[i],vibrR[i]+j);//vibrato rate
                                    midiOutShortMsg(lphmo[i],vibrD[i]+j);//vibrato depth
                                    midiOutShortMsg(lphmo[i],0x005b00+0xb0+j);//reverb set to 0
                        }
//---------Sounds initialization (16×8)--------------------------------------------------------------------
for( int i = 0; i < 8; i++)
                        for( int j = 0; j < 16; j++){
                                    midiOutShortMsg(lphmo[i],0x7f0000+notes[j]+0x90+j);
                                    setVol(0,i,j); // Volume is set to 0 waiting for the first image
                        }
```

Finally, for each pixel, the following messages are sent inside setVol, depending on the depth of the pixel (parameter “dept”), its horizontal position (kept in “pan”) and its vertical position (transmitted in the channel “ch” variable, as seen in figure 6.57):

```
if( dept < 3 )
        dept = 2;
midiOutShortMsg(lphmo[pan],((dept/2)<<16)+0x0700+0xb0+ch);            //Channel volume
if( dept < 50 )
        dept = 50;
midiOutShortMsg(lphmo[pan],(((dept-50)/2)<<16)+0x4a00+0xb0+ch);       //CutOff frequency
```

6.5. Quantitative Evaluation

In this section the sonification will only be quantitatively evaluated, i.e., regarding technical aspects.

The processing time is computed for the 16×8 pixels image used in the previous examples.

Profile	Time	
	Processing image	Sonification
P0	87.0 µs (11472 fps)	3.3µs (298295 fps)
P1	12.8 µs (77816 fps)	3.6 µs (275349 fps)
P2	11.0 µs (89488 fps)	4.9 µs (238636 fps)
P3	10.0 µs (99431 fps)	5.3 µs (188397 fps)
P4	12.8 µs (77816 fps)	10.6 µs (94198 fps)
P5	9.5 µs (105280 fps)	17.3 µs (57734 fps)
P6	9.2 µs (108472 fps)	22.6 µs (44191 fps)

Table 6.11. Time results of each profile.

6.6. Evaluation

The tests designed to validate the sonification protocol were based over static images and a Virtual Reality (VR), to provide deeper information about the usability and accuracy of the sonification protocol.

6.6.1. User Tests

The sonification protocol was tested with a set of 28 individuals. 17 undergraduates and postgraduates students from the Georgia Institute of Technology, plus 11 from the Center for the Visually Impaired (CVI) of Atlanta (Georgia) participated in this study. The sample was composed by 11 males and 17 females, with a mean age of 33.46 years, range 18-62. Among them, 13 were sighted, 10 presented low vision and 5 were completely blind. All reported normal or correct to normal hearing. The group presented the following demographic characteristics:

- Sighted group, exclusively students from the GeorgiaTech, an average age of 21.51 years old (range between 18 and 26), educational level of 3.35 (see Survey in Annex III; the range is a 5 levels Likert scale) and computer use of 4.94 (the same scale applies).
- Visually impaired group, mostly from the Center for the Visual Impaired of Atlanta (CVI), aged 51.91 years old with a range between 36 and 64 years old, with an average educational level of 2.7 and a computer use of 3.82.

Table 6.11 shows the correlations between the four demographic variables:

- AGE: the age in years the day of the test.
- VI: Visual Impairment with three ordered values: sighted (1), low vision (2) and blind (3).
- EDU: Educational level with four values: elementary school (1), high school (2), some college (3) and college degree or higher (4).
- COMP: Use of computer with five ordered values about the use of computers: never (1), rarely (2), once a week (3), once a day (4) and many times a day (5).

Pearson correlation	**AGE**	**VI**	**EDU**	**COMP**
AGE	1	.752** (p<.001)	-0.458* (p=0.014)	-0.708** (p<.001)
VI		1	-0.474* (p=0.011)	.637** (p<.001)
EDU			1	-0.527* (p=.004)
COMP				1

Table 6.12. Pearson's correlation index and p-value. * marks significant correlations at a level of .05 and ** marks significant correlations at a level of .001.

As it can be seen, parameters that should be independent were found to be highly correlated in the testing groups, which won't allow analyzing them separately. Use of computer, age and visual impairment are highly correlated and, thus, the analysis will be done over the use of computer, since it is the most descriptive variable (the age is quite variable and the visual impairment is so narrow with only 3 different values).

6.6.1.1. Experiments

The tests were carried out in the Sonification Lab, at the School of Psychology of the Georgia Institute of Technology, under the supervision of Prof. Bruce Walker and the Institutional Review Board of the Office of Research Compliance [103] of the same Institute.

In the Annex III it is detailed the experiment set up, the survey and control forms and some adjacent information related to this experimentation.

Nevertheless, in the following lines a summary of such experiments is given. A virtual reality (VR) environment was developed with the Unity3D engine [104], connected through the IServer program to the InterSense InertiaCube2 head tracker [105]. The Unity3D rendered images are sent to the sonification program, written in C and connected to the V-Stack [101] and Edirol HQ Hyper Canvas Synth [99]. This synthesizer allows processing General MIDI 2 signals (see Annex I) produced by the sonification program, whose correlative sounds are transmitted to the user by means of a pair of earphones. Objective data, about correct and incorrect perceptions and decisions was gathered during the tests. Finally, subjective data was gathered after the test with a survey the participants had to fulfill.

The experiment was thought to validate two different uses of the system: as a mobility aid (its main purpose) and as artificial vision system. Thus, a set of experiments was designed to validate these two approaches.

The complete set of steps in the experiment was designed as follows:

- VR training: Participants were briefly trained in one single level, which was assigned to them randomly (but according to their visual status (divided as sighted, low vision and blind), to cover all the cases as uniformly as possible). This training was done over static images (available over the online test of the protocol at http://163.117.201.122/validacion_ATAD_cerrada) for around 5 minutes. After this step, they passed through 7 training scenes (figure 1) that were designed in Unity3D.

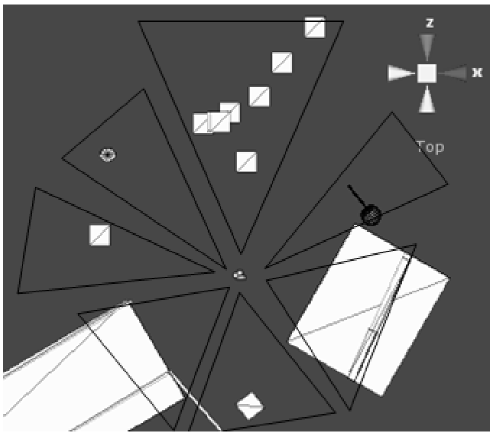


Fig. 6.70. Unity3D training environment with the 7 scenes surrounding the user, static, in the center.

The VR training consisted of 7 scenes with different objects, some of them static and some others performing periodic movements in different axis. The participant had the opportunity of facing (clockwise order from the upper cone in figure 6.70) a stack of boxes, a pendulum, an open door, a box moving horizontally near a wall, a corridor, a box moving like an infinite symbol closer and farther from the participant and a column. When they had difficulties to see the scenes being sonified, they were verbally described by the experimenter. In every case, they couldn't see the screen after the training and the only feedback of the virtual reality was provided through the earphones. This step had no time limit, although all of the participants completed it in less than 20 minutes.

- VR artificial vision test 1: Four testing scenes were also designed in Unity3D, and they are shown in figure 6.71.

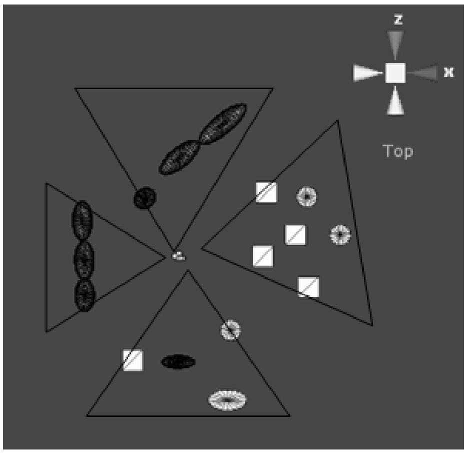


Fig. 6.71. Four testing scenes around the participant avatar, located in the center.

Starting from the left one, the scenes were composed of:

- Scene 1: three balls at three different heights, same distance (located in a 3×3 grid from the user point of view).

- Scene 2: three balls at three different heights, three different distance (located in a 3x3 grid from the user point of view) and the farther one repetitively moving from the bottom height to the middle height and back to the bottom.
- Scene 3: 6 boxes and balls at three heights and distances (located in a 3×3 grid from the user point of view).
- Scene 4: four objects in the same horizontal line, the left one slightly changing its position in sudden movements every 10 seconds.

In the first three setups, the participants were asked to guess where were the objects in the 3×3 grid (up, middle, down and left, center and right). They didn't know the number of objects that were present in each test.

- VR mobility tests 2 and 3: The participant is left in two labyrinths and receives the instructions of following the wall at their right all the time, and as fast as possible. The forward and backward movements were done manually in the keyboard of the computer in fixed steps each time the participant said "go forward" or "go backward". The direction change was done rolling on a chair and transmitted by the head tracker. In the test 2, the users were asked to follow the wall at their right hand as fast as possible, without missing the reference. If they said "lost", they were manually taken to the last position to keep performing the task, and this fact was written down. This labyrinth can be completed in 42 steps. The second labyrinth (test 3) present a similar structure, but some objects have been placed, some of them taller than the avatar controlled by the user (such as columns), some others shorter. In this second case, the users are asked to follow the wall, to report each time they think they are close to an object and to say "lost" whenever they don't know where is the wall. This labyrinth can be completed in 50 steps. Both labyrinths are shown in figure 6.72.

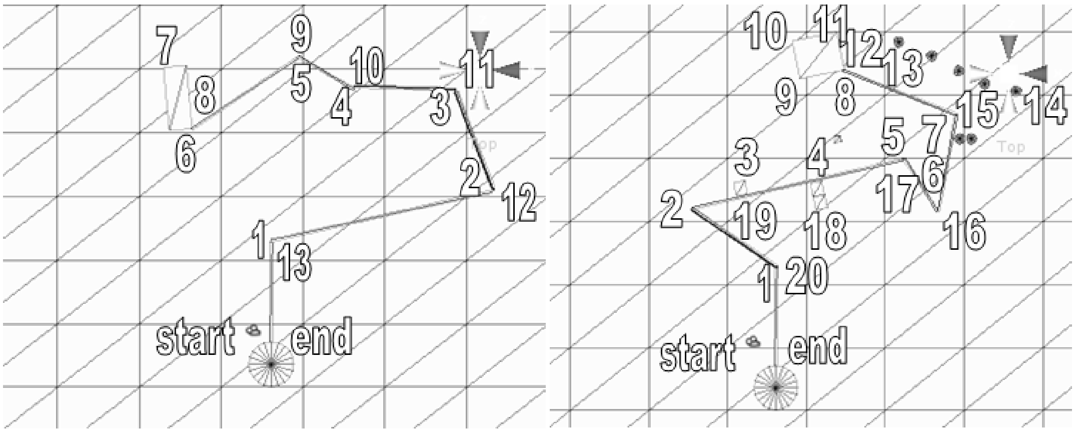


Fig. 6.72. Top view of the labyrinths of test 1 (left) and test2 (right).

As it can be seen, the labyrinths are divided in milestones which serve as check points. A time limit of 10 minutes was given to the users to perform the complete task. The obstacles in the second test were a column attached to the wall (milestone 3), a column at the left (4), a closed corner (6), a corridor of columns (13-15), a lower obstacle (18) and a second closed corner (19).

- VR combined test 4: The forth test was done over a room, where the users were left free to move in any direction, with the instructions of reporting each time they perceived something, and describe it in terms of “low” or “high” obstacle. They had 20 minutes to virtually walk around the room. Sighted participants were also asked to draw a map of the room, with the found objects on it, marked as “L” (for lower obstacles) and “H” (for higher ones). The room is shown in figure 6.73.

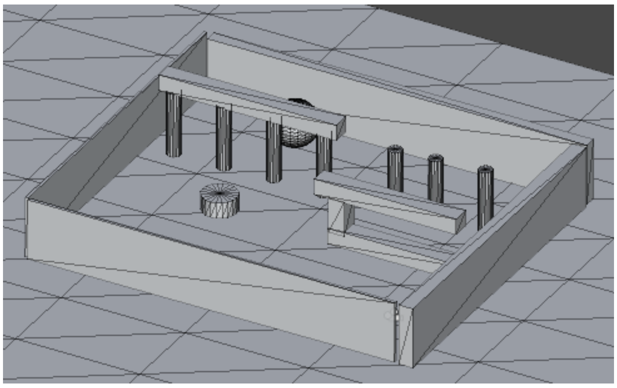


Fig. 6.73. Perspective view of the virtual room (test 3). The starting point is the lower corner in the figure.

- Subjective perceptions: As said, participants were asked to fulfill a survey at the end of the process. They were asked about subjective perceptions about the training and the test, as well as a general evaluation of the process. When the subjective reports from participants are linked to some specific task or test, they will be presented in the corresponding subsection. A final subsection about the general evaluation will be provided apart.

6.6.1.2. Results

The results will be organizing following the tests described above.

Test 1 Results

Regarding the scene 1 of the test 1, the participants answered guessing the position of the objects, correctly identifying an average of 1.71 of them (s.d. of .854) and a false positives average of 1.39 (s.d. of 1.197). No significant correlations were found when comparing these results with the educational level, the use of computer, the age or the profile level. However, significant differences were found when comparing the number of false positives with the visual impairment (one way ANOVA: F(2,25)=3.728, p=.028). Figure 6.74 shows this result.

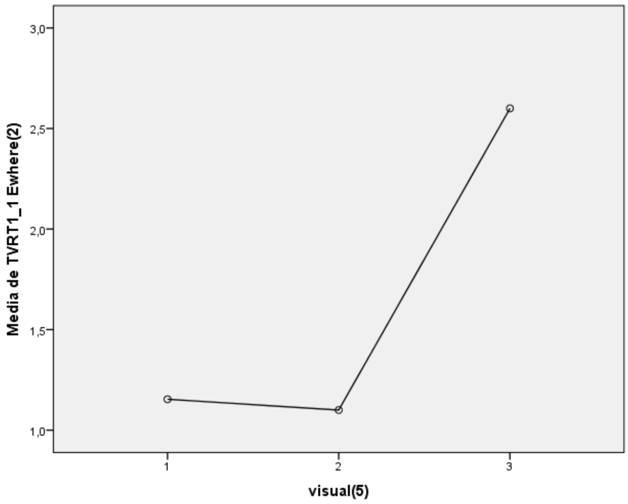


Fig. 6.74. False positives against visual impairment.

Although the difference of means between errors and correct detections and educational level was not significant (F(2,25)=.648, p=.532 for the false positives, F(2,25)=1.311, p=.287 for the correct detections), their representation remains interesting, as shown in figure 6.75. This result will be discussed later.

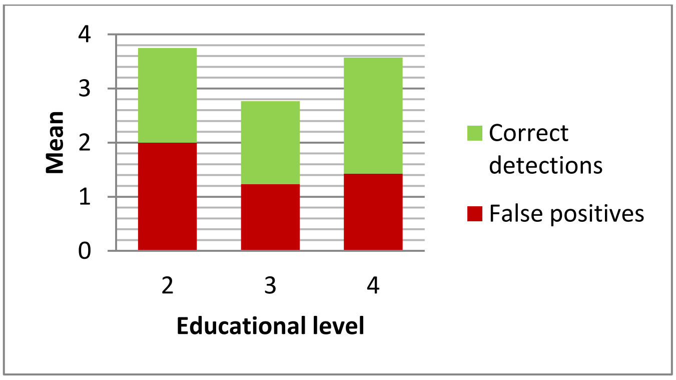


Fig. 6.75. Mean of correct detections (in green) and false positive (in red) against educational level.

In the second scene, the number of correctly detected objects presented a mean of 1.29 (s.d. .763) and the same mean for the false positives (with s.d. of 1.197). No significant differences were found comparing these two variables with the different factors already discussed. However, the number of false positives in this test and that of the previous one presented a significant Pearson correlation (.497, p=.007). Although not statistically significant, there can be seen a relation between the number of errors and the profile level used, as shown in figure 6.76.

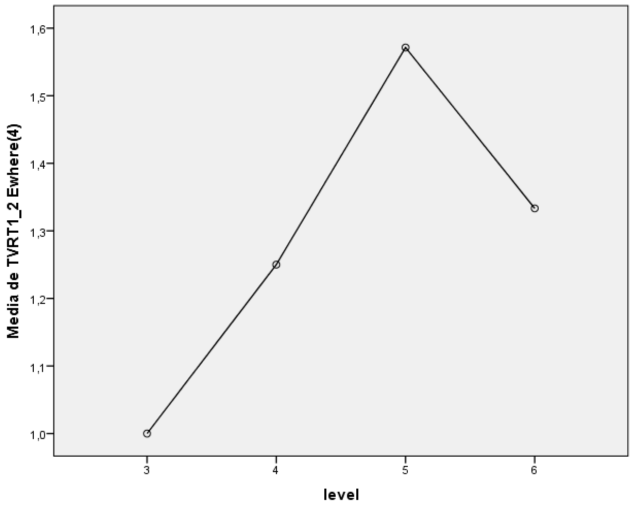


Fig. 6.76. False positives in test 2 against the profile level.

The correct detections don’t follow the same rule.

Likewise, the mean of the false positives and correct detections against the visual impairment also provides interesting data, as shown in figure 6.77.

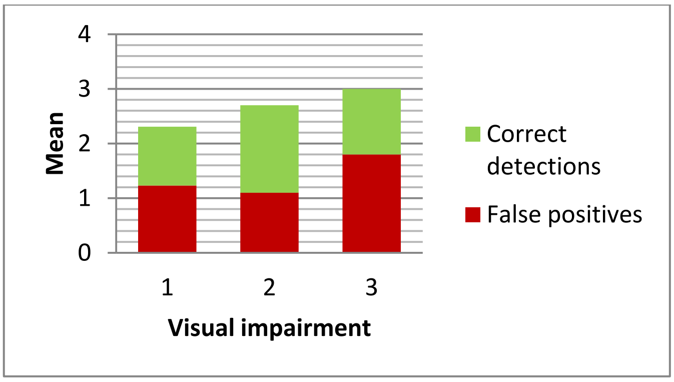


Fig. 6.77. Correct detections (in green) and false positives (in red) against visual impairment.

There is another evident relation (although not significant) between the correctly detected objects and the use of computer, as shown in figure 6.78.

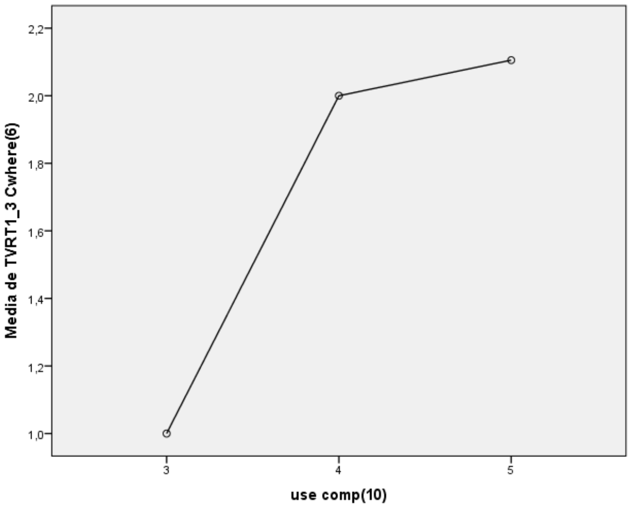


Fig. 6.78. Mean of the correctly detected objects in test 2 against the use of computers. (For this figure, the only sample with use of computer level of 2 has been eliminated).

The third scene produced different results. On the one hand, the average number of errors was .5 (s.d. .745) and the mean of the correctly detected objects was 1.96 (s.d. 1.17). Given that there were 6 objects, only the 32.6% of the objects were detected.

Marginally significant Pearson correlations (.328, p=.088) was found when correlating the correct detections and the profile level. The result of the one way ANOVA analysis was not significant (F(3,24)=1.906, p=.156) even though a tendency can be easily seen:

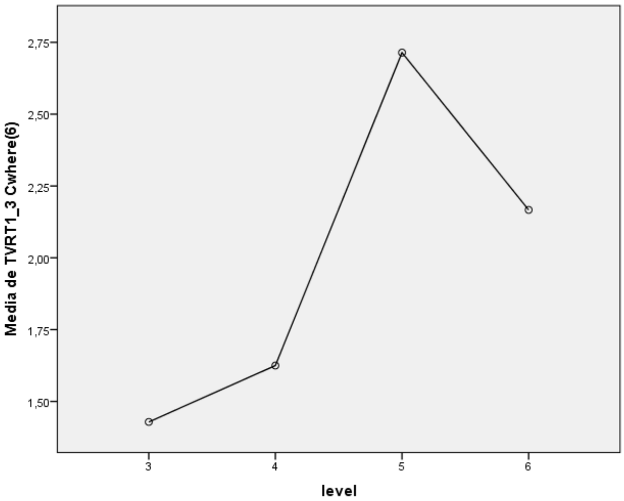


Fig. 6.79. Correct detections in test 3 against the profile level.

Once again, the low vision group presented the best results, as seen in figure 6.80.

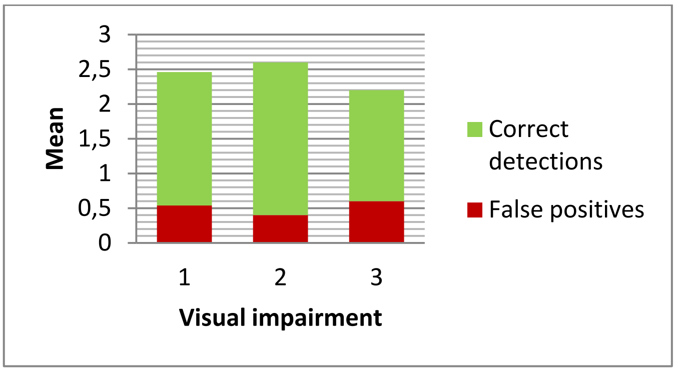


Fig. 6.80. Mean of correct detections (in green) and false positives (in red) in test 3 against the visual impairment.

In the forth scene, the average number of correctly detected objects in the forth scene was 1.79 (s.d. 1.134) and that of false positives .79 (s.d. 1.101). Only two participants reported to have found more than 3 objects. Once again, no significant results were found when calculating the Pearson correlations between pairs of variables and factors.

Other relevant results found were the mean of the number of detected objects in terms of the profile level can be seen in figure 6.81.

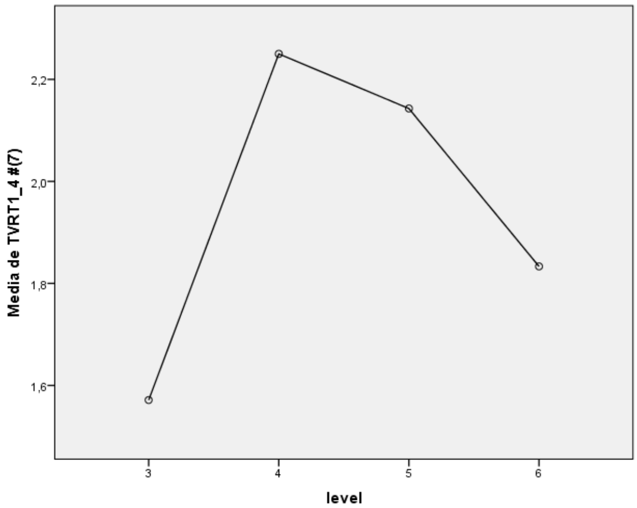


Fig. 6.81. Mean of the number of correctly detected objects in terms of the profile level.

Not significant but presenting a clear rule the relation between the same variable and the educational level can be appreciated in figure 6.82.

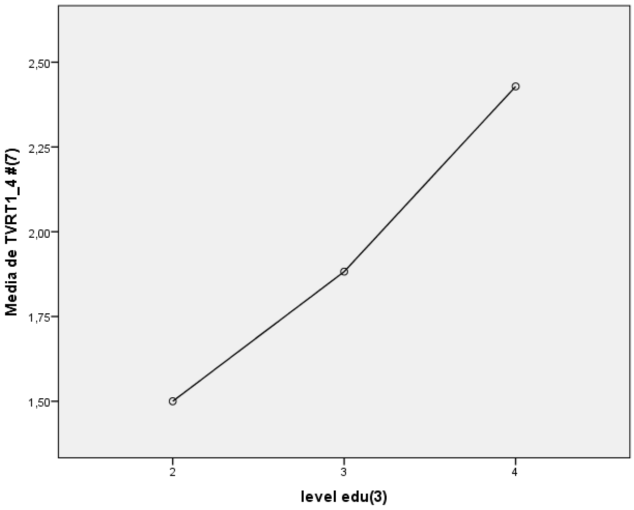


Fig. 6.82. Mean of the number of correctly detected objects against the educational level.

In this test participants were asked to identify the moving object in the horizontal plane. The mean of correct identification index (ranged 0 –no detection- and 1 –correct detection) of the moving object was .79 (.69 in the sighted, .9 in the low vision and .8 in the blind group). The average number of objects correctly ordered in terms of distance was 1.79 (1.62 for the sighted, 2 for the low vision and 1.8 for the blind groups).

Test 2 Results

First of all, we found that the 67.8% of the users achieved the end point and the 85.7% crossed the middle point in less than 10 minutes, both groups in less than 10 minutes. Highly significant results were found with a one-way ANOVA mean comparison between the use of computer and the position achieved after the 10 minutes (F(3,24)=7.265, p=0.001, shown in figure 6.83).

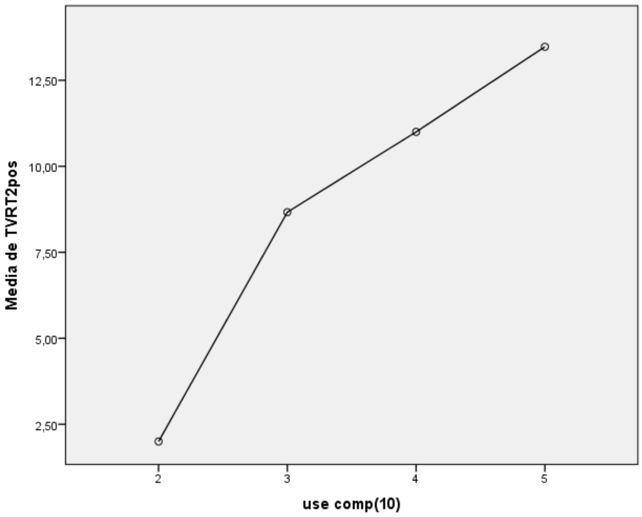


Fig. 6.83. Final position of the avatar (from 1 to 14) after 10 minutes or less compared to the use of computer.

Similar results were found when comparing the position against the visual impairment (F(2,25)=9.239, p=.001), the position against the educational level (F(2,25)=10.755, p<.001), and the time to finish against the visual impairment (F(2,25)=4.146, p=.028).

Another interesting result was found when comparing the time required to finish (if less than 10 minutes) against the profile. Significant results were found with a one way ANOVA analysis (F(3,24)=4.691, p=.01). However, in the figure 6.84, we see an interesting change of shape.

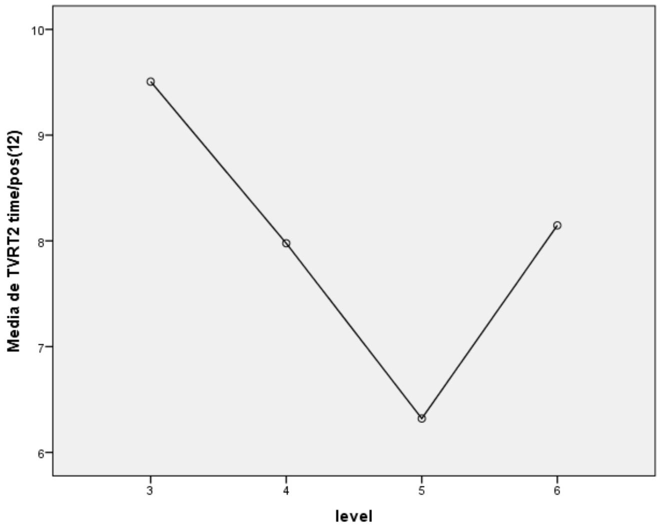


Fig. 6.84. Time (in minutes) required to finish compared to the profile.

Looking at the subjective perception of this test, the perception of ease of achieving the task was inversely correlated (but not statistically significant) with the use of computer, as seen in figure 6.85 (Pearson non-significant correlation of -.276, p=.155).

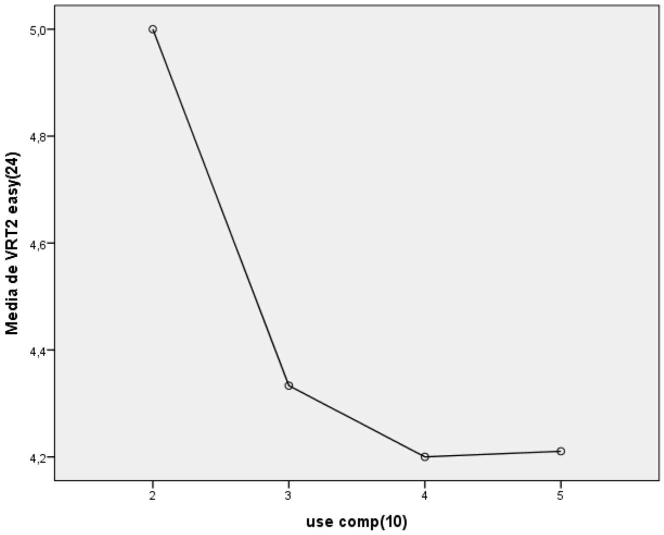


Fig. 6.85. Ease perceived in test 1 in terms of use of computer.

Similar relations were found when comparing the means of the perception of ease and the educational level (F(2,25)=2.247, p=.147).

When asked about the difficulty presented by corners, significant results were found with a one way ANOVA with the profile level as factor (F(3,24)=6.365, p=.005). Not statistically significant by very interesting result was the relation between the ease perceived and the profile level, shown in figure 6.86.

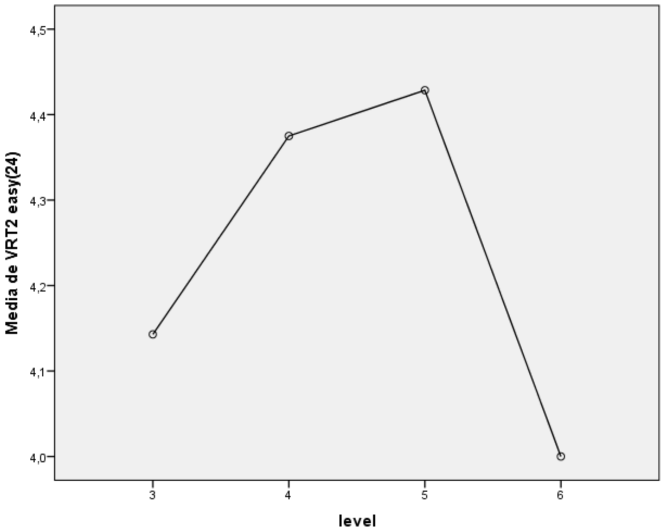


Fig. 6.86. Ease perceived against profile level.

Test 3 Results

In the third test, some more variables were measured: how many obstacles were detected and which ones.

Regarding the objective data, 32.1% of the participants achieved the end point in less than 10 minutes, and the 78.5% crossed the middle point in less than 10 minutes.

No significant results were found when comparing the achieved position after 10 minutes or less with the use of computer (ANOVA, F(3,24)=1.037, p=.394) or the profile level (ANOVA, F(3,24)=.613, p=.613). However, in this last case, we find a similar distribution than that shown in figure 6.87.

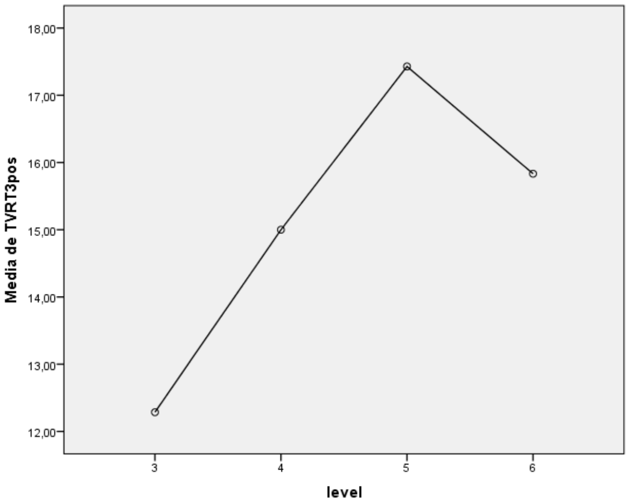


Fig. 6.87. Achieved position (ranged between 1 and 21) in terms of profile level.

The descriptors of correct detection of the different obstacles are shown in table 6.13.

Obstacle	Mean	Std. Dev
Attached column	.11	.315
Column	.39	.49
Corner1	.22	.42
Columns	.56	.51
LowBlock	.59	.51
Corner2	.50	.53

Table 6.13. Obstacles detections.

The correct detection of obstacles also depended of the profile level following a similar pattern:

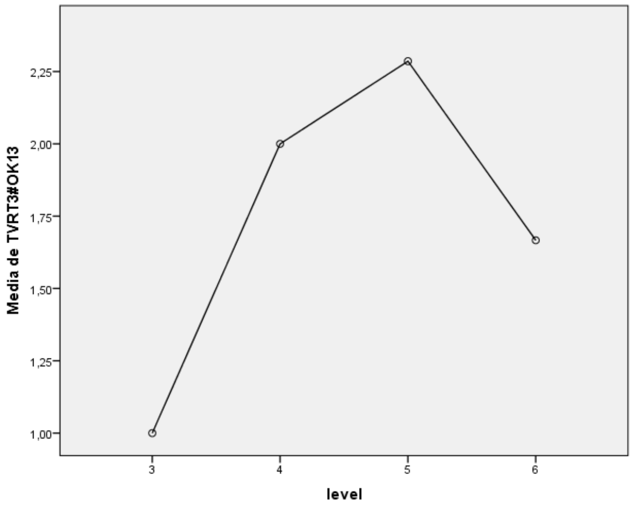


Fig. 6.88. Number of correctly detected obstacles in terms of the profile level.

These differences were, however, not significant (F(3,24)=1.687, p=.196).

Another interesting result was the distribution of probability of detection of the lower block in terms of the profile level, shown in figure 6.89.

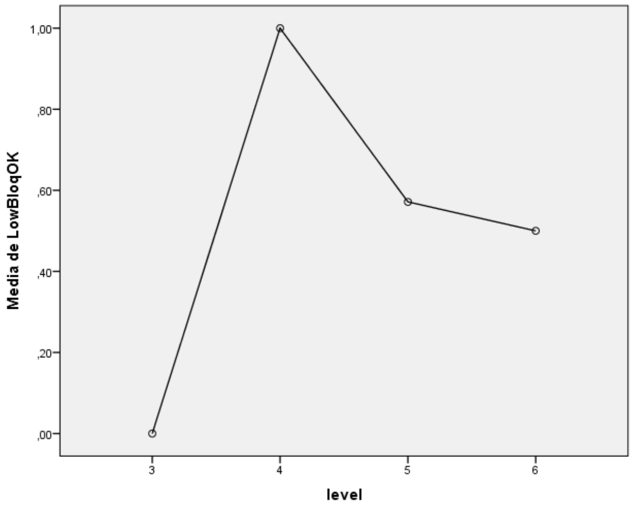


Fig. 6.89. Detection level (ranged between 0 –no detection- and 1 –detected-) of the lower block in terms of profile level.

The one way ANOVA means comparison was not found to be significant (F(3,13)=2.240, p=.132).

We also found statistically significant differences between the number of correctly detected obstacles and the visual impairment (F(2,25)=4.366, p=.024), shown in figure 6.90.

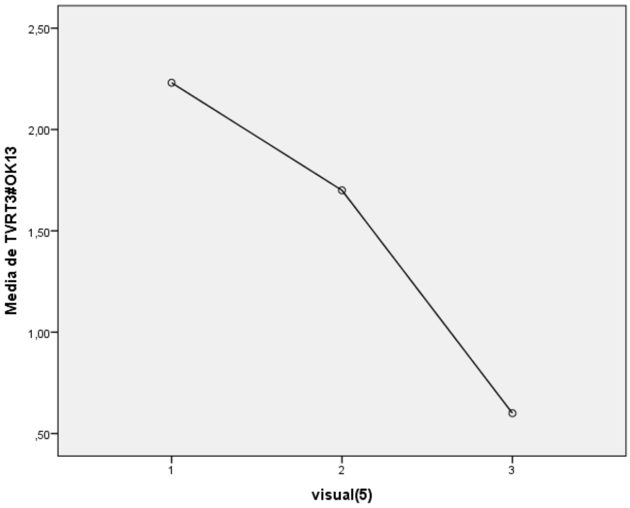


Fig. 6.90. Number of correctly detected obstacles in terms of the visual impairment.

Attending to the subjective results, no significant results were found with profile level, use of computer and visual impairment as factors, when the participants were asked for ease of detecting bigger or lower obstacles.

Once again, users perceived the ease of detecting obstacles differently in terms of the profile level, as shown in figure 6.91.

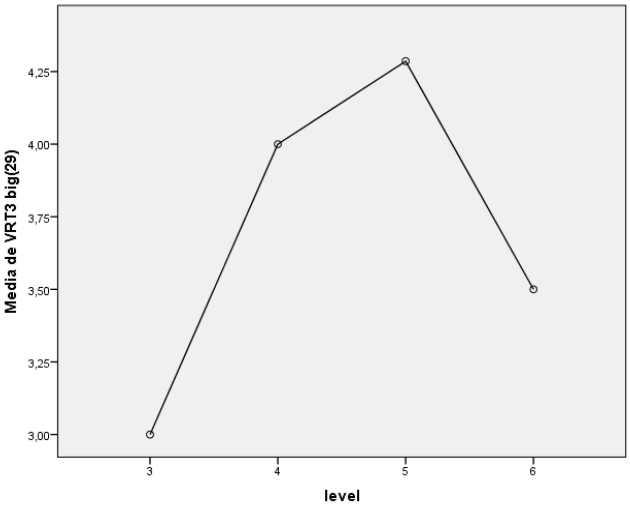


Fig. 6.91. Ease of perception of bigger obstacles in terms of the profile level used.

This difference was not significant (F(3,24)=1.755, p=.183).

Test 4 Results

The room test provided some significant results. The Pearson correlation and one way ANOVA analysis showed relevant correlations between detection of the right set of columns and the visual impairment (Pearson correlation = -.599, p=.001. ANOVA: F(2,24)=6.778, p=.005), the sphere detection and the same parameter (Pearson correlation = -.466, p=.014. ANOVA: F(2,24)=4.689, p=.019) and between the use of computer and the detection of the sphere (Pearson correlation =.501, p=.008. ANOVA: F(3.23)=2.627, p=.075, marginally significance).

When analyzing the subjective perception of the test, although none of the results were significant, some interesting relations can be appreciated. On the one hand, the blind people had to put more effort to produce a mental image of the room, as it can be seen in figure 6.92.

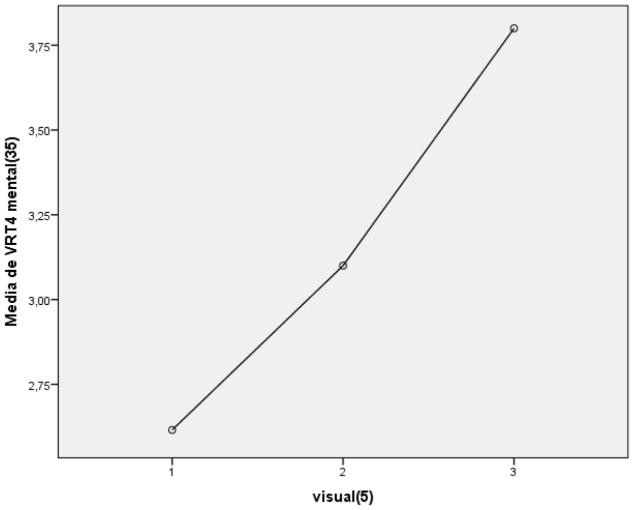


Fig. 6.92. Mental effort (ranged from 1 –no effort- to 5 –high effort-) against visual impairment.

In a similar way, the perception of being lost in the room was also related with the same factor, as shown in figure 6.93.

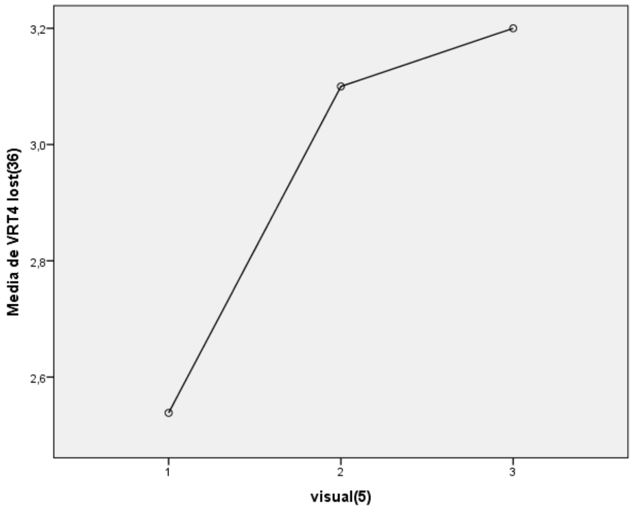


Fig. 6.93. Sensation of being lost (ranged from 1 –not at all- to 5 –every time-) against the visual impairment.

Final Evaluation Results

Final questions about the global process were asked to the participants, among which the tiredness of the global process, considerations about the length of the training, feelings of safeness and use of the white cane or the guide dog.

Statistically significant Pearson correlation was found between the educational level and the perception of the tiredness of the whole process (-.384, p=.044). This result is shown in figure 6.94.

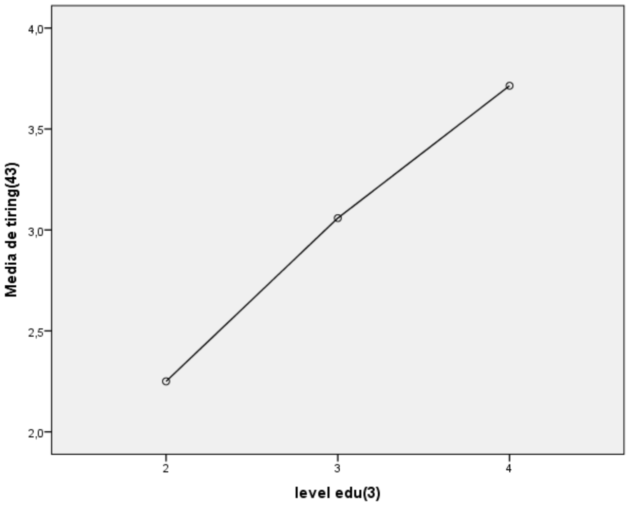


Fig. 6.94. Perception of tiredness (ranged between 1 –not tiring at all- and 5 –very tiring-) against the educational level.

Another interesting result is the distribution of the same perception against the level, shown in figure 6.95.

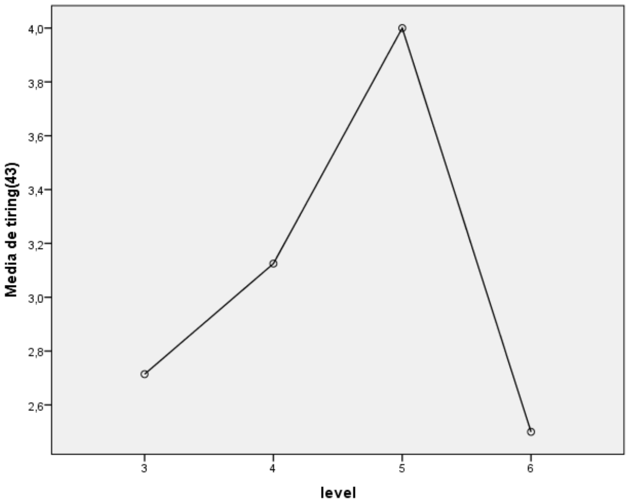


Fig. 6.95. Perception of tiredness (ranged between 1 –not tiring at all- and 5 –very tiring-) against the profile level.

The final relation found (not significant, one way ANOVA F(2,25)=.014, p=.986) between perception of tiredness and the visual impairment is shown in figure 6.96.

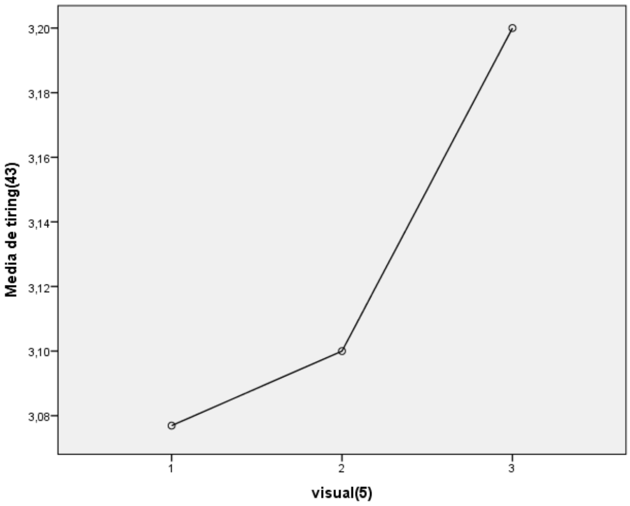


Fig. 6.96. Tiredness perception against visual impairment.

An unexpected result comes from the comparison of the use of computer and the perception of safeness (marginally significant with a one way ANOVA, F(3,24)=2.707, p=.068), shown in figure 6.97.

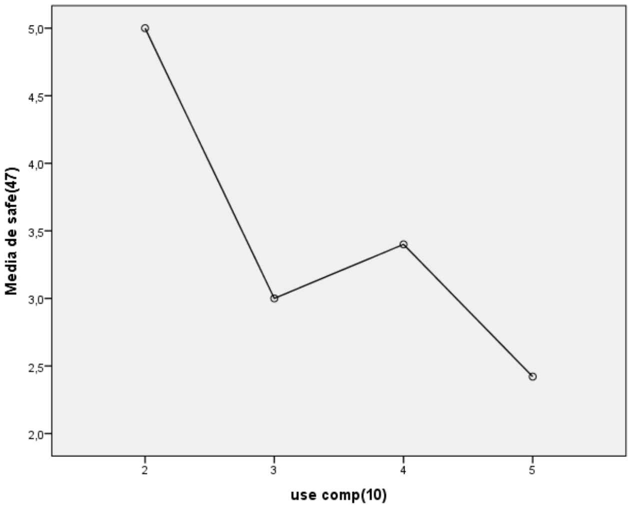


Fig. 6.97. Perception of safeness (ranged between 1 –not safe at all- and 5 –very safe-) against the use of computer.

Another marginally significant result (one way ANOVA, F(2,11)=3.378, p=.072) is the relation between the intention of keep using the white cane or the dog guide against the visual impairment (with a positive Pearson correlation of .585 and p=.028).

The lower correlation was found between the visual impairment and the perception of tiredness (.030, p=.878).

6.7. Discussion

First of all, we have to discuss the specific composition of the participants' pool available for this experiment, as well as its size. The highly correlation found between visual impairment and other theoretically independent variables such as use of computer, educational level or age is due to the composition of the clients and staff of the CVI, compared to the average student of the GeorgiaTech, much younger and used to use computers daily.

Regarding the second point, 28 participants are not a very big set of subjects, and the quantitative data obtained should be contrasted with larger experiments. This one can be taken as a preliminary study of the usability and efficiency of the system proposed in [106].

The sonification had its own conditions for a proper functioning regarding the user requirements, being the first of them the real time constraint. This worked in two main ways:

- The sonification itself must be adaptable to a frame-to-frame sonification paradigm.
- The sonification must be light enough to allow cheap hardware cores to generate it live.

The first condition affects the design. We had presented the available parameters to generate a sonification in table 6.6. In that table, we saw that not every parameter is suitable for a frame-to-frame sonification, i.e. each frame must produce a complete set of sounds, independently of the time it lasts. However, we allowed one exception to this rule: the vibrato. It was introduced because the GM2 MIDI implementation allowed managing it as a frame-to-frame parameter (we set the vibrato level at the boot, and it is always working independently of the loudness of the sound to which it is applied). The rest of the sonification parameters were chosen respecting this paradigm. We can summarize them in the following list:

- Vertical axis divided between 1 and 8 -or 16 in the software version- levels (depending on the cognitive level), assigning to each one different tones composing a harmonic chord.
- Horizontal axis divided in 1 or 8 (depending on the cognitive level), below the capacity of horizontal discrimination of the human hearing system.
- Vibrato level proportional to the distance to the middle columns.
- Distance mapped into loudness and number of harmonics, producing strident sounds when the object is near.

Regarding the second condition, in table 6.9 are shown the computational differences between different approaches in generating complex sounds. We rejected the multiband filtering because of both memory and, over all, computational load, and looked forward the pregenerated and the on-the-flight generation. This last option was finally chosen, because of software implementation ease, with both MIDI and puredata standards. The MIDI option, in the software version (PC with 1.6GHz CPU), was found to be very fast: 22.6µs (44191fps).

Another condition was the need of an intuitive sonification. This was, again, implemented in two ways:

- A mixed sonification paradigm, as defined in {Revuelta Sanz, 2013 444 /id}, which exploits the psychoacoustic habitudes, introducing as less arbitrary paradigms as possible.

- The organization of the sonification in different cognitive levels, to allow each user using the one that gives her/him more information with not that much effort. When s/he is ready to change the level, the new one will only complete the previous and, hence, all the previous training will be useful.

Regarding the first issue, we recovered some results in the user tests, describing the level of difficulty of understanding scenes through sounds. In figure 6.86 we see that the average level of ease perceived with the sonification, for every cognitive level, is over 4 in a 5 points Likert scale. In figure 6.92, we can see how the average level of the mental effort needed to understand the *room test* was under the neutral value for sighted and low vision participants, although higher in the case of blind participants. The ease of localizing objects in each user centered axis, horizontal, vertical and distance, presented a mean value of 4.3 (s.d. of .5), 4.5 (.9) and 3.7 (1.3) respectively. This was paradoxical, since the psychoacoustic based mappings were found to be harder to be understood. Anyhow, these results demonstrated that the sonification mapping was easy to use.

When attending to the second way, we found interesting results correlating the ease perceived against the level used, demonstrating, as shown in figure 6.86, a positive correlation between the ease and the level, with the exception of level 6, which dramatically decreases the ease of perception. This point was also remarked by the experts during the focus group. This fact invited us to eliminate this last level in the hardware implementation. Moreover, we could propose a relation between the profile level and the efficiencies of the tests (and, maybe, the perception of usability by the users), which was tested independently of the demographic characteristics of the participants' pool, assigning levels equally to each visual category. In the tests 2, 3 and 4, we find important changes of efficiency and subjective perceptions in terms of the profile level assigned. When comparing the time required to finish (if less than 10 minutes) against the profile in the first test, significant results were found (F(3,24)=4.691 and p=.01, figure 6.84). The results of the third test (figures 6.87 and 6.88) were found to be not significant but with the same shape. Significant or not, the performance in different aspects reach a maximum with the profile 5, with the exception of the detection of the lower block in the second test. This exception can be due to specific differences between users. It is important to notice that not every user achieved this block (57.1% of them faced the lower block, one of them with profile level 3, with profile level 4, 6 with profile level 5, 4 with profile level 6). We find two apparently contradictory results, shown in figures 6.76 and 6.79. In the first one, the average number of false positives (in the second scene) increases until level 5, and then decreases for the sixth. In the second one, the number of correct detections (in the third scene) increases until the level 5, and then decreases for the sixth level. This second result seems to be easier to be explained, since the higher is the profile level, the more information is available for the user, with the restriction of the understanding of complex combinations of objects, which could lead to a decrease of efficiency in the last level. In the second case, when evaluating the efficiency of the sonification in the forth scene, a simple interpretation can be done: since the height in this test was not relevant (and this was an instruction for the participants), and only horizontal and distance information had to be provided, the higher is the profile level, the more redundant information is being provided for the user and, thus, the more problems s/he may encounter to understand the horizontal position of the objects. In the first case, that of errors in the first scene of the test 1) we can explain this result: it was difficult to understand the scene when the same object produced more than one tone. In that case (and remembering that we are representing false positives), the higher is the level seems to be related (although not statistically significance) with the problems in understanding the number of objects. This is not contradictory with the precedent argument, because higher levels allow understanding easier the vertical combination of objects (with the marked limits) but, at the same time, may produce the perception of "phantom objects" added to the real ones. Anyhow, profile level 5 seems to be the most efficient and, at the same time, is perceived as so (figure 6.76 relating the errors in the VR scene 1 test, figure 6.81 with the correct detections in the forth scene of the test 1, figure 6.86 for test 2, figure 6.92 for test 3 and figure 6.95 for general evaluation of the tiredness of the process).

Many evidences to support the general hypothesis that other factors modulates the usability and efficiency of the system were found (see figures 6.74, 6.75, 6.77, 6.78, 6.80, 6.82, 6.83, 6.85, 6.90, 6.92 and 6.93 about objective results, 6.94 and 6.96 about subjective perceptions, with their correlative statistic data), however, and due to the specific demographic composition of the subjects pool, we cannot find significant and specific differences between the COMP, AGE, EDU and VI factors. Even if age, use of computer or visual impairment are somehow correlated in our dataset, sometimes we can find significant results focusing our attention in some specific factor. In all these cases, the higher is the familiarity with the computers (and the higher is the educational level), the better are the results. In the first case, figures 6.78 is coherent with the idea that using daily the computer helps when using this kind of systems. However, among these results we find unexpected relations. Let's consider the fact that the lower is the use of computers makes people think the completion of the labyrinth was easier (figure 6.85 and correlative statistical data). This does not match with the objective data about completing it (F(3,24)=7.265, p=0.001, and shown in figure 6.83), and some conjectures can be proposed:

- There can be differences between the time perception of the people used to manage computers and/or doing exams and those who are not.
- The level of self-exigency could be modulated by the educational level.
- There can be differences in the conservatism and the notion of risk (and, thus, the difficulty to perform a task correctly) between young and older people.

In the second case (attending to the educational level), figure 6.75 shows how lower educational level leads to more risky (more trials and hence more correct and incorrect results), a more conservative approach in the first years of college and, finally, the higher rate of correct over incorrect results. Figure 6.82 shows a clear (but not significant) relation between educational level and the number of detected objects. This is also coherent with the idea that a higher educational level also helps. When evaluating the effect of the visual impairment, we found unexpected results. Data represented in figures 6.74, 6.77 and 6.80 shows the low vision group as the best one when interpreting the sonification protocol, and the hypothesis of an ordered decrease of performances from sighted to blind people is not justified. We could relate this with the effort needed to understand the scene (as also shown in figure 6.96).

A boolean alarm, requested by the experts interviewed, as shown in chapter 3, section 3.3.3, was also implemented, as cognitive level 0, although it was not tested in real situations cause no participant chose this level.

Our hypothesis H3 states that the sonification is useful to represent spatial configurations of objects. We found in the VRE tests that the number of correct identification of objects normally was higher than the incorrect one (1.96 versus .5 in the third scene, 1.79 versus .79 in the forth, 1.71 versus 1.39 in the first and the same in the second). It can be concluded, with some restrictions, that the sonification protocol proposed describes, with a short training, the real world. The restrictions are related with the design of the sonification, the hearing limits in the perception and the influence of other variables, as will be discussed now.

The correlative hypothesis states that the sonification helps from a mobility point of view. We found that the 67.8% of the users achieved the end point in the first labyrinth and the 85.7% crossed the middle point in less than 10 minutes, both groups in less than 10 minutes, with a training of 15 minutes. Given that the sonification protocol presents some aspects which are not intuitive at all, this results can be taken as an evidence of the correctness of our assumption. The same hypothesis could be checked in the second test, although the efficiency decreased in the second labyrinth. This can be due to the following factors:

- The second labyrinth was a 20% longer than the first one, but the time available remained the same.
- The second labyrinth presented some obstacles, which had to be found (according to verbal instructions to the users), and many users spent some time exploring their environment to decide whether some sounds represented an object or just the empty ground or the wall.

The amount of obstacles found is not so high, a mean of 2.93 and a standard deviation of 1.3. Given that there were 6 tagged obstacles (two of them narrow corners which were ignored by some participants), slightly less than the half of them were perceived as obstacles. More in detail, the column attached to the wall was only perceived as an obstacle by 3 participants. The rest of them simply dodged it as if it was the normal shape of the path. The limit of what can be considered as an obstacle and what is part of the background is not always clear. The low block and the columns were detected by the half of the participants that reached those points. According to these results, perceiving a flat wall seems to be easier than some objects.

Regarding the design itself of the sonification, and the limits in the hearing system, we can appreciate the important reduction in the detection rate of objects between the first and second scene of the test 1. As seen in figure 6.71, the first scene presents objects at the same distance (or almost), while the second one present the objects at three different distances from the observer. This fact dramatically reduced the correct detection rate in the second scene, because of the limits in the auditory perception of the hearing system.

Regarding the general evaluations, we can find again unexpected results, such as the direct correlation between educational level and the feeling of tiredness after the tests. People with higher educational level should be more prepared for intellectual and sometimes boring tasks, but they manifested the higher rates of tiredness (Pearson correlation of -.384 and p=.044, figure 6.94). The comparison of the safeness feeling with the use of computer gives us another unexpected result (F(3,24)=2.707 and p=.068, figure 6.97): the group with better results (with higher rate of computer use) felt more unsafe than that with lower performances. The reasons can be related to those pointed out when discussing about the perception of ease of the labyrinth test.

Finally, as expected, blind people want to keep using the white cane or the guide dog even if they could use this system (4.8 over 5 as average response against 3 of the sighted group and 4.33 of the low vision group). Given that blind people use to encounter harder dangers in the middle height than in the bottom part [107], and the comment of the experts about the higher ease of detecting middle and higher obstacles, the system can increase the safeness in the travels of this collective.

The main limitations of this study (the demographic composition of the participant subgroups and the size of the sample) should be the first tasks to be performed in the future, with the goal of obtaining statistically stronger data to describe the use, efficiency and usability of the system.

Longer training should be tested to evaluate the real potentials of the system, and more real life tests should also be designed.

6.8. Conclusions

Our hypothesis H2 about the capability of the sonification to help visually impaired people to perceive information about the spatial configuration of their surroundings has been confirmed. However, this translation must be trained.

In this study we have tested a new sonification protocol over a set of 28 people, sighted, with low vision and blind, trying to find its limitations and strengths.

We found that different combinations and structures of sounds can bring to differences in the efficiency and even in the usability of a sonification-based assistive product.

We also found relations with other aspects not specifically related with disabilities, such as educational level, use of computer in the daily life or the age.

Some simplifications can be done on the system (such as eliminating the most complex profile level, or “cleaning” the presented information to give the user only the most relevant one).

The most efficient user of the tested assistive product is a young person, used to manage computers and with some college or a college degree educational level. People not matching these characteristics seem to face harder problems to understand or at least take the right decisions in mobility tests with this system. However, they use to feel safer and they find the system less tiring.

The research carried out in this field, and explained in this chapter, gave two peer reviewed publications in conference proceedings and journal, as shown in the following references:

- Authors: Pablo Revuelta, Belén Ruiz Mezcua, José Manuel Sánchez Pena, Bruce N. Walker.

Title: Scenes and images into sounds: a taxonomy of image sonification methods for mobility applications
Journal: Journal of the Audio Engineering Society (in press)
Date: March 2013.

- Authors: Pablo Revuelta Sanz, Belén Ruiz Mezcua, José M. Sánchez Pena
Title: A Sonification Proposal for Safe Travels of Blind People.
Conference: The 18th International Conference on Auditory Display (ICAD2012).
Publication: (not available yet).
Place: Atlanta, Georgia (U.S.A.).
Date: June 18-22, 2012

References

1. M. Capp and Ph. Picton, "The Optophone: An Electronic Blind Aid." *Engineering Science and education Journal* vol. 9 no. 2, pp. 137-143. 2000.

2. J. A. Ramírez Rábago, "Generación de fuentes virtuales de sonido en audífonos.,", Escuela de Ingeniería, Universidad de las Américas, Puebla, Mayo., 2005.

3. F. Miraya, *Acústica y Sistemas de Sonido*: UNR Editora, 2006.

4. J. Gil Soto, "Psico-Acústica." Centro de Innovación de Automoción de Navarra, ed. 2008.

5. A. Rodríguez, "* (revisar ref.type) Conceptos Básicos de la Psicoacústica, Seminario de Audio." . 2005. Uruguay, Instituto de Ingeniería Eléctrica IIE Facultad de Ingeniería - UDELAR Montevideo.

6. E. H. Weber, *De Pulsu, Resorpitone, Auditu et Tactu: Annotationes Anatomicae et Physiologicae.*, Leipzig, Germany: Koehlor, 1834.

7. G. T. Fechner, *Elements of Psychophysics.*, New York, 1966.

8. ISO, *ISO 532, Acoustics - Method for calculating loudness level.*, Geneva: ISO, 1975.

9. H. Fletcher and W. A. Munson, "Loudness, its definition, measurement and calculation," *Journal of the Acoustical Society of America,* vol. 5, no. 2. pp.82-108, 1933.

10. ISO, "ISO 226 (R2003), Acoustics-Normal equal-loudness level contours. Geneva." *http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=34222* . 2003. ISO.

11. B. N. Walker and M. A. Nees, "Theory of Sonification." *Handbook of Sonification (in press)*. A. H. &. J. N. E. In T.Hermann, ed. 2012. New York, Academic Press.

12. S. S. Stevens and J. Volkmann, "The relation of pitch to frequency: A revised scale.," *The American Journal of Psychology,* vol. 8. pp.329-353, 1940.

13. B. Kapralos and M. R. M. Jenkin, "Auditory perception and spatial (3d) auditory systems." vol. CS-2003-07. July 2003.

14. W. G. Gardner, *3D Audio and Acoustic Enviroment Modelling*: Wave Arts Inc., 1999.

15. J. Blauert, *Spatial hearing: the psychophysics of human sound localization*: MIT Press Cambidge, 1983.

16. L. Rayleigh, "On our perception of sound direction," *Philos. Mag.,* vol. 13. pp.214-232, 1907.

17. W. L. Hartman, *How we localize sounds*: AIP Press, 1997.

18. W. A. Yost, "Lateral position of sinusoids presented with interaural intensive and temporal differences." *Journal of the Acoustical Society of America* vol. 70 no. 2, pp. 397-409. 1981.

19. G. F. Kuhn, "Model for the interaural differences in the azimuthal plane." *Journal of the Acoustical Society of America* vol. 62, pp. 157-167. 1977.

20. G. Peris-Fajarnés, I. Dunai, and V. Santiago Praderas, "Detección de obstáculos mediante sonidos acústicos virtuales." *DRT4ALL 2011*. *DRT4ALL 2011* , pp. 133-139. 2011.

21. C. Kyriakakis, "* (revisar ref.type) Fundamental and technological limitations of immersive audio systems." *Proceedings of the IEEE*. vol. 86. 1988.

22. W. M. Hartmann and A. Wittenberg, "On the externalization of sound images." *Journal of the Acoustical Society of America* vol. 99 no. 6, pp. 3678-3688. 1996.

23. D. S. Brungart, "Auditory localization of nearby sources III.Stimulus effects." *Journal of the Acoustical Society of America* vol. 106 no. 6, pp. 3589-3602. 1999.

24. D. J. Kistler and F. L. Wightman, "Resolution of front-back ambiguity in spatial hearing by listener and source movement." *Journal of the Acoustical Society of America* vol. 105 no. 5, pp. 2841-2853. 1999.

25. M. Pec, M. Bujacz, P. Strumillo et al., "Individual HRTF Measurements for Accurate Obstacle Sonification in an Electronic Travel Aid for The Blind." *International Conference on Signals and Electronic Systems (ICSES 2008)* , pp. 235-238. 2008.

26. F. F. Romanow, "Methods for Measuring the Performance of Hearing Aids," *The Journal of the Acoustical Society of America,* vol. 13, no. January, 1942.

27. J. P. Dolch, "Phase and Intensity Relationships in the Interference of Bone-Conducted and Air-Conducted Sound," *Journal of the Acoustical Society of America,* vol. 26, no. 5. pp.942, 1954.

28. N. Watson and T. V. Frazier, "Equal Loudness Contours for Hearing by Bone Conduction," *Journal of the Acoustical Society of America,* vol. 24, no. 1. pp.114, 1952.

29. R. W. Carlisle, H. A. Pearson, and P. R. Werner, "Construction and Calibration of An Improved Bone-Conduction Receiver for Audiometry," *Journal of the Acoustical Society of America,* vol. 19, no. 4. pp.632-638, 1947.

30. N. A. Watson, "Hearing Aids: Uniform and Selective: Monaural, Diotic and Binaural; Air and Bone Conduction," *The Journal of the Acoustical Society of America,* vol. 19, no. 4, 2012.

31. E. P. Fowler, "Fundamentals of bone conduction," *Archives of Otolaryngology,* vol. 2, no. 6. pp.529-542, 1925.

32. S. Stenfelt and R. L. Goode, "Transmission properties of bone conducted sound: Measurements in cadaver heads," *Journal of the Acoustical Society of America,* vol. 118, no. 4. pp.2373-2391, 2005.

33. G. C. Saleeby, G. D. Allen, R. B. Mahaffey et al., "Air Conduction and Bone Conduction = Your Own Voice," *Journal of the Acoustical Society of America,* vol. 59. pp.S15, 1976.

34. S. P. Y. Stenfelt and B. E. V. Hakansson, "Sensitivity to bone-conducted sound: excitation of the mastoid vs the teeth," *Scandinavian Audiology,* vol. 28, no. 3. pp.190-198, 1999.

35. G. Flottorp and S. Solberg, "Mechanical Impedance of Human Headbones (Forehead and Mastoid Portion of Temporal Bone) Measured Under Iso-Iec Conditions," *Journal of the Acoustical Society of America,* vol. 59, no. 4. pp.899-906, 1976.

36. S. Stenfelt, B. Hakansson, and A. Tjellstrom, "Vibration characteristics of bone conducted sound in vitro," *Journal of the Acoustical Society of America,* vol. 107, no. 1. pp.422-431, 2000.

37. C. E. Dean, "Audition by bone conduction," *Journal of the Acoustical Society of America,* vol. 2, no. 2. pp.281-296, 1930.

38. P. Carlsson, B. Hakansson, and A. Ringdahl, "Force Threshold for Hearing by Direct Bone-Conduction," *Journal of the Acoustical Society of America,* vol. 97, no. 2. pp.1124-1129, 1995.

39. M. Anjanappa, X. Chen, R. E. Bogacki et al., "Method and Apparatus for Tooth Bone Conduction Microphone." no. 7,486,798. 1982.

40. T. Koulis, J. O. Ramsay, and D. J. Levitin, "From zero to sixty: Calibrating real-time responses," *Psychometrika,* vol. 73, no. 2. pp.321-339, 2008.

41. W. J. Dowling, "Music, cognition, and computerized sound: An introduction to psychoacoustics," *Contemporary Psychology-Apa Review of Books,* vol. 47, no. 1. pp.36-38, 2002.

42. C. Ramakrishnan, "Sonification and Information Theory." *CMMR/ICAD 2009, LNCS* vol. 5954, pp. 121-142. 2010.

43. C. Yeh, A. Röbel, and X. Rodet, "Multiple Fundamental Frequency Estimation of Polyphonic Music Signals." *ICASSP 2006* , p.III-225-III-228. 2005.

44. W. Starkiewicz and T. Kuliszewski, "The 80-channel elektroftalm." *Proceedings of the lnternational Congress Technology Blindness, Am.Found.Blindness* vol. 1, p.157. 1963. nEW yORK.

45. L. W. Farmer, "Mobility devices," *Bull Prosthet Res*. pp.47-118, 1978.

46. F. d'Albe, *The moon element*, London: T. Fisher Unwin, Ltd., 1924.

47. R. L. Beurle, *Summary of suggestions on sensory devices*, London: San Dunstan's, 1947.

48. E. Milios, B. Kapralos, A. Kopinska et al., "Sonification of range information for 3-D space perception." *IEEE Transactions on Neural Systems and Rehabilitation Engineering* vol. 11 no. 4, pp. 416-421. 2003.

49. L. W. Farmer, "Mobility devices," *Bull Prosthet Res*. pp.47-118, 1978.

50. Jie X, W. Xiaochi, and F. Zhigang, "Research and Implementation of Blind Sidewalk Detection in Portable ETA System." *International Forum on Information Technology and Applications* , pp. 431-434. 2010.

51. S. Shoval, J. Borestein, and Y. Koren, "Auditory Guidance with the Navbelt-A Computerized Travel Aid for the Blind." *IEEE Transactions on Systems, Man, and Cybernetics* vol. 28 no. 3, pp. 459-467. 1998.

52. A. Fusiello, A. Panuccio, V. Murino et al., "A Multimodal Electronic Travel Aid Device." *Proceedings of the Fourth IEEE International Conference on Multimodal Interfaces* , pp. 39-44. 2002.

53. F. Fontana, A. Fusiello, M. Gobbi et al., "A Cross-Modal Electronic Travel Aid Device." *Mobile HCI 2002, Lecture Notes on Computer Science* vol. 2411, pp. 393-397. 2002.

54. T. Ifukube, T. Sasaki, and C. Peng, "A Blind Mobility Aid Modeled After Echolocation of Bats," *IEEE Transactions on Biomedical Engineering,* vol. 38, no. 5. pp.461-465, 1991.

55. J. Gonzalez-Mora, A. Rodriguez-Hernandez, E. Burunat et al., "Seeing the world by hearing: Virtual Acoustic Space (VAS) a new space perception system for blind people," *International Conference on Information & Communication Technologies: from Theory to Applications (IEEE Cat.No.06EX1220C)*. pp.6-ROM, 2006.

56. Y. Kawai and F. Tomita, "A Support System for Visually Impaired Persons Using Acoustic Interface - Recognition of 3-D Spatial Information." *HCI International* vol. 1, pp. 203-207. 2001.

57. Y. Kawai and F. Tomita, "A Visual Support System for Visually Impaired Persons Using Acoustic Interface." *IAPR Workshop on Machine Vision Applications (MVA 2000)* , pp. 379-382. 2000.

58. Y. Kawai and F. Tomita, "A Support System for Visually Impaired Persons Using Three-Dimensional Virtual Sound." *Int.Conf.Computers Helping People with Special Needs (ICCHP 2000)* , pp. 327-334. 2000.

59. M. M. Fernández Tomás, G. Peris-Fajarnés, L. Dunai et al., "Convolution application in environment sonification for Blind people." vol. VIII Jornadas de Matemática Aplicada, UPV. 2007.

60. N. Ortigosa Araque, L. Dunai, F. Rossetti et al., "Sound Map Generation for a Prototype Blind Mobility System Using Multiple Sensors." *ABLETECH 08 Conference* , p.10. 2008.

61. P. B. L. Meijer, "An Experimental System for Auditory Image Representations," *IEEE Transactions on Biomedical Engineering,* vol. 39, no. 2. pp.112-121, 1992.

62. L. H. Riley, G. M. Weil, and A. Y. Cohen, "Evaluation of the Sonic Mobility Aid." vol. American Center for Research in Blindness and Rehabilitation, pp. 125-170. 1966.

63. G. Sainarayanan, R. Nagarajan, and S. Yaacob, "Fuzzy image processing scheme for autonomous navigation of human blind." *Applied Soft Computing* vol. 7 no. 1, pp. 257-264. 2007.

64. A. D. Heyes, "The use of musical scales to represent distance to object in an electronic travel aid for the blind." *Perceptual and Motor Skills* vol. 51 no. 2, pp. 68-75. 1981.

65. A. D. Heyes, "Human Navigation by Sound." *Physics in Technology* vol. 14 no. 2, pp. 68-75. 1983.

66. A. D. Heyes, "The Sonic Pathfinder - A new travel aid for the blind." *In Technology aids for the disabled*. W.J.Perk and Ed. s, eds. pp. 165-171. 1983. Butterworth.

67. A. D. Heyes, "The Sonic Pathfinder - A new travel aid for the blind." *In Technology aids for the disabled*. W.J.Perk and Ed. s, eds. pp. 165-171. 1983. Butterworth.

68. A. D. Heyes and G. Clarcke, "The role of training in the use of the Sonic Pathfinder." *Proceedings of the American Association for the Education and rehabilitation of the Blind and Visually Impaired, Southwest Regional Conference, Hawaii.* 1991.

69. L. Kay, "Auditory perception of objects by blind persons, using a bioacoustic high resolution air sonar," *Journal of the Acoustical Society of America,* vol. 107, no. 6. pp.3266-3275, 2000.

70. N. C. Darling, G. L. Goodrich, and J. K. Wiley, "A preliminary followup study of electronic travel aid users," *Bull Prosthet Res,* vol. 10, no. 27. pp.82-91, 1977.

71. R. F. A. Rossi, M. K. Zuffo, and J. A. Zuffo, "Improving spatial perception through sound field simulation in VR," *2005 IEEE International Conference on Virtual Environments, Human-Computer Interfaces and Measurement Systems (IEEE Cat.No.05EX1045C)*. pp.103-108, 2005.

72. M. Bujacz, P. Skulimowski, G. Wróblewski et al., "A proposed method for sonification of 3D environments using scene segmentation and personalized spatial audio." *Conference and Workshop on Assistive Technologies for People with Vision and Hearing Impairments Past Successes and Future Challenges*. *CVHI 2009* , pp. 1-6. 2009.

73. F. Zhigang and L. Ting, "Audification-based Electronic Travel Aid System." *IEEE International Conference on Computer Design and Applications (ICCDA 2010)* vol. 5, pp. 137-141. 2010.

74. D. A. Ross and B. B. Blasch, "Wearable Interfaces for Orientation and Wayfinding." *ASSETS'00* , pp. 193-200. 2000.

75. D. Castro Toledo, S. Morillas, T. Magal et al., "3D Environment Representation through Acoustic Images. Auditory Learning in Multimedia Systems." *Proceedings of Concurrent Developments in Technology-Assisted Education* , pp. 735-740. 2006.

76. D. Aguerrevere, M. Choudhury, and A. Barreto, "Portable 3D sound / sonar navigation system for blind individuals." *2nd LACCEI Int.Latin Amer.Caribbean Conf.Eng.Technol.* pp. 2-4. 2004.

77. P. Baranski, M. Polanczyk, and P. Strumillo, "A remote guidance system for the blind," *2010 12th IEEE International Conference on e-Health Networking, Applications and Services (Healthcom 2010)*. pp.386-390, 2010.

78. B. N. Walker and G. Kramer, "Human factors and the acoustic ecology: Considerations for multimedia audio design." *Proceedings of the Audio Engineering Society 101st Convention*. *Proceedings of the Audio Engineering Society 101st Convention* . 1996. Los Angeles.

79. A. R. O'Hea, "Optophone design: optical-to-auditory substitution for the blind,", The Open University, U.K., 1994.

80. R. M. Fish, "Audio Display for Blind," *IEEE Transactions on Biomedical Engineering,* vol. 23, no. 2. pp.144-154, 1976.

81. G. Balakrishnan, G. Sainarayanan, R. Nagarajan et al., "Fuzzy matching scheme for stereo vision based electronic travel aid," *Tencon 2005 - 2005 IEEE Region 10 Conference, Vols 1-5*. pp.1142-1145, 2006.

82. S. A. Dallas and A. L. Erickson, "Sound pattern generator representing matrix data format|has matrix video converted to parallel form, modulating audio tone, giving video information in terms of time and frequency." THALES RESOURCES INC, ed. no. WO8200395-A1; EP55762-A; US4378569-A; CA1165447-A; IL63239-A; EP55762-B; DE3174174-G. 1960.

83. M. A. Hersh and M. Johnson, "Mobility: An Overview." *Assistive Technology for Visually Impaired and Blind People*. Marion A.Hersh and Michael A.Johnson, eds. no. 5, pp. 167-208. 2008.

84. BESTPLUTON World Cie, "The "Mini-Radar", your small precious companion that warns you obstacles in a spoken way, and that helps you to walk straight." *http://bestpluton.free.fr/EnglishMiniRadar.htm* . Apr. 2011.

85. G. Balakrishnan, G. Sainarayanan, R. Nagarajan et al., "Stereo Image to Stereo Sound Methods for Vision Based ETA." *1st International Conference on Computers, Communications and Signal Processing with Special Track on Biomedical Engineering, CCSP 2005, Kuala Lumpur* , pp. 193-196. 2005.

86. L. Kay, "KASPA." *http://www.batforblind.co.nz/* . 2005.

87. A. Mittal and S. Sofat, "Sensors and Displays for Electronic Travel Aids: A Survey." *International Journal of Image Processing* vol. 5 no. 1, pp. 1-14. 2010.

88. G. Balakrishnan, G. Sainarayanan, R. Nagarajan et al., "Fuzzy matching scheme for stereo vision based electronic travel aid," *Tencon 2005 - 2005 IEEE Region 10 Conference, Vols 1-5*. pp.1142-1145, 2006.

89. H. G. Kaper, E. Wiebel, and S. Tipei, "Data Sonification and Sound Visualization." *Computing in Science & Engineering* vol. July/August 1999, pp. 48-58. 1999.

90. L. M. Brown, S. A. Brewster, R. Ramloll et al., "Design guidelines for audio presentation of graphs and tables." *Proceedings of the International Conference on Auditory Display (ICAD2003)* , pp. 284-287. 2003. Boston.

91. R. Cusack and B. Roberts, "Effects of differences in timbre on sequential grouping," *Perception & Psychophysics,* vol. 62, no. 5. pp.1112-1120, 2000.

92. B. N. Walker and G. Kramer, "Ecological psychoacoustics and auditory displays." http://sonify.psych.gatech.edu/publications/ . 2004.

93. T. W. Payne, "Working memory capacity and pitch discrimination,", Georgia Institute of Technology, Atlanta, 2003.

94. C. S. Watson and G. R. Kidd, "Factors in the design of effective auditory displays." *Proceedings of the International Conference on Auditory Display (ICAD1994)*. 1994. Santta Fe.

95. P. B. L. Meijer, "An Experimental System for Auditory Image Representations," *IEEE Transactions on Biomedical Engineering,* vol. 39, no. 2. pp.112-121, 1992.

96. MIDI Manufacturers Association MMA, "General MIDI 1, 2 and Lite Specifications." http://www.midi.org/techspecs/gm.php . 2012. 23-9-2011.

97. S. Jordà Puig, *Audio digital y MIDI,* Guias Monográficas, Madrid: Anaya Multimedia, 1997.

98. Microsoft., "Microsoft MIDI Mapper." http://msdn.microsoft.com/en-us/library/windows/desktop/dd798702%28v=vs.85%29.aspx . 2011.

99. Roland, "The Edirol HyperCanvas HQ GM2." http://www.roland.com/products/en/HQ-GM2/ . 2012.

100. Wikipedia, "Virtual Studio Technology." *en.wikipedia.org/wiki/Virtual_Studio_Technology* . 2011.

101. Steinberg, "V-Stack." http://www.kvraudio.com/product/v_stack_by_steinberg . 2011.

102. MidiOX, "MIDI Yoke." http://www.midiox.com/ . 2011.

103. Office of Research Compliance, "Institutional Review Board." http://www.compliance.gatech.edu/ . 2012.

104. Unity, "Unity3D." http://unity3d.com/ . 2012.

105. ISense, "InertiaCube head Tracker." http://www.intersense.com/pages/18/11/ . 2010.

106. P. Revuelta Sanz, B. Ruiz Mezcua, and J. M. Sánchez Pena, "A Sonification Proposal for Safe Travels of Blind People." *Proceedings of the 18th International Conference on Auditory Display (ICAD 2012)* , pp. 233-234. 2012. Atlanta, GA.

107. P. Revuelta Sanz, B. Ruiz Mezcua, and J. M. Sánchez Pena, "Users and Experts Regarding Orientation and Mobility Assistive Technology for the Blinds: a Sight from the Other Side." *Proceedings of the AIRTech Int.Conference 2011.* 2011.

7. System Integration and Evaluation

The different parts developed until now respond to the interfaces proposed in chapter 4. This proposal specified some requirements to be implemented in each step of the signal processing chain. In this chapter, we present the system architecture as well as the interfaces among the different subsystems in order to obtain a single functional system.

The integration of both the image processing and the sonification subsystems will be done in two steps: (i) a software integration, which will serve as preliminary validation of the whole system, by means of specific user tests. In this section, we will also include a training proposal to ease the user's doing the tests; (ii) after that, the description of the hardware implementation will be exposed and the device tested in real-life environments by potential users.

Please keep in mind the basic architecture presented in chapter 4, more specifically in the figure 4.4.

7.1. Software Integration

The software preliminary implementation is mandatory in a flexible and adaptive design of any system. This task allows us to check if what we are producing fits with the users' demands and requirements. Moreover, flexibility in software designs makes possible changes with minimum cost and, finally, a more accurate design. In this section we will show how the different information is managed all over the whole processing chain.

In every case, the image processing algorithm will be the full definition version (all lines scanned) of the last algorithm, the fast and dense dynamic processing one.

7.1.1. Description

As it was shown in chapter 4, fig. 4.4, the system consists of two image sensors, an image processing step, a sonification step and a transmission device.

7.1.1.1. System Set-Up and Programming Environments and Languages

In the software integration, we will use as complementary HW two low-cost USB webcams [1] with a resolution of 320×240 pixels at 30 fps, and around 90^{o} of field of view. These cameras are attached to a laptop.

The software environment is the Microsoft Visual Studio 2005, with the image processing library OpenCV, which is free and open source. All the programs are written in ANSI C.

However, this library has only been used to capture images from the cams and to show them. All the functions and filters have been implemented in a matrix level and, thus, will not be used in the hardware implementation. Sonification part, as said, was written in MIDI format. The real-time MIDI messages are produced following the MIDI tables available in annex I. The MIDI synthesizer is called by means of the WinMM.dll library of Windows. These functions can be called from C programs, by means of the following functions:

```
WINMMAPI MMRESULT WINAPI midiOutOpen( OUT LPHMIDIOUT phmo, IN UINT
uDeviceID, IN DWORD_PTR dwCallback, IN DWORD_PTR dwInstance, IN DWORD fdwOpen);
WINMMAPI MMRESULT WINAPI midiOutClose( IN OUT HMIDIOUT hmo);
WINMMAPI MMRESULT WINAPI midiOutShortMsg( IN HMIDIOUT hmo, IN DWORD dwMsg);
```

7.1.1.2. Information Flowchart

The figure 7.1 shows the information flow through the different stages of the processing chain.

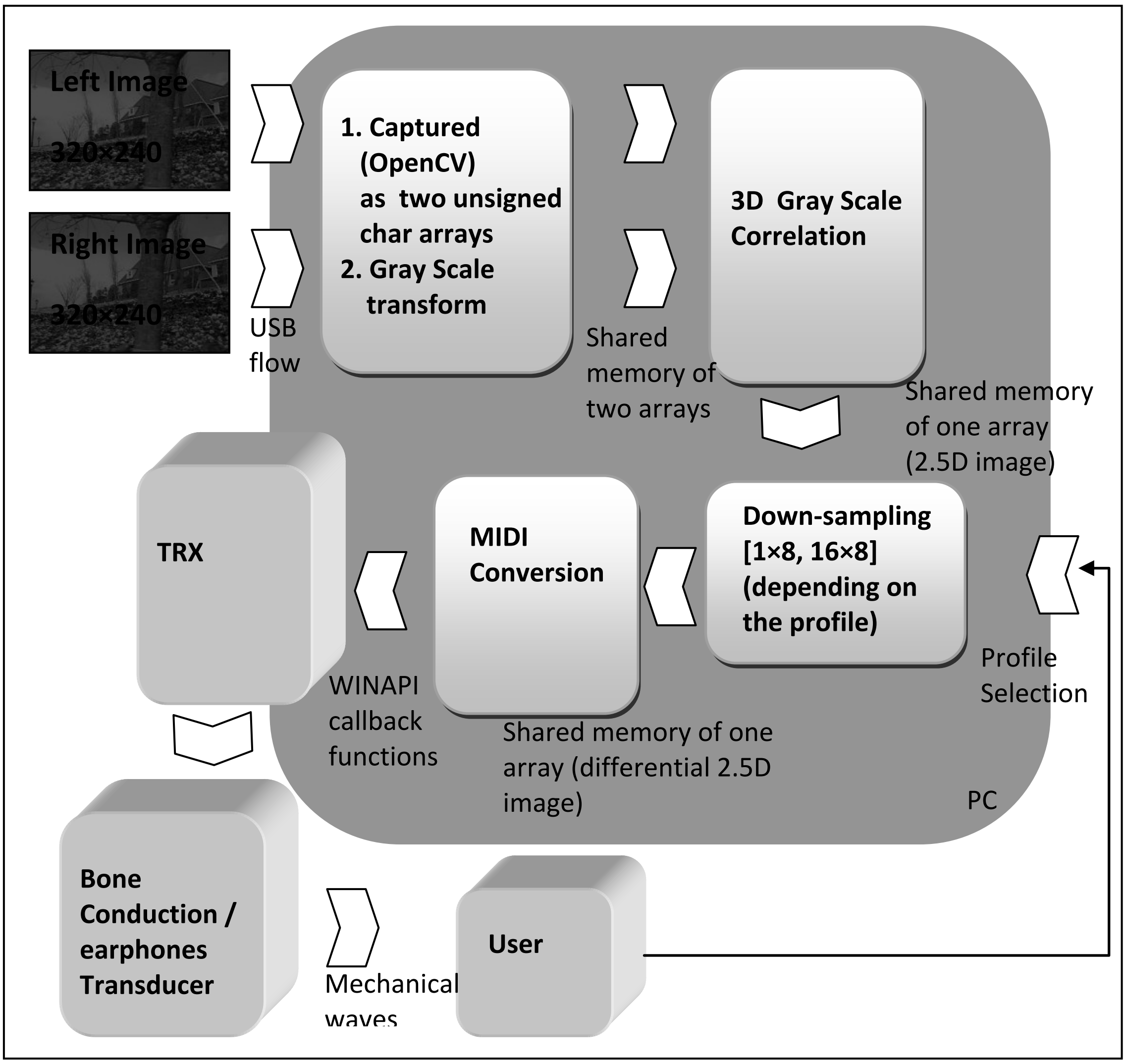


Fig. 7.1. Flowchart of the software implementation of the system. Protocols and instances involved and the information path.

7.1.2. Quantitative Evaluation

A first evaluation on the performance of the program can be made in terms of frames per second. No data can be obtained related with errors of the stereo vision algorithm, since we have no depth truth image to compare.

The processing rate of the system, processing images at 320×240 and sonifying, running on a 1.6GHz single core processor, and 1GB of RAM memory, is around 32 fps (ranged 20 to 38) measured over 15 images at different illumination conditions.

7.1.3. User Tests

The whole system in its software version was assembled and validated at the Sonification Lab, School of Psychology of the GeorgiaTech, Atlanta, GA, U.S.A., during the summer of 2012, and with the collaboration of the Center for the Visually Impaired of Atlanta.

The evaluation was designed in two main steps, one with a larger sample of participants, to obtain quantitative data, and another one with a smaller amount of participants but with longer and deeper training and testing, as well a focus group at the end to obtain qualitative data about the experience and the system itself.

The composition of the larger sample of participants is described in section A.II.2. The demographics of the experts group are detailed in A.III.2.

7.1.3.1. Experiments Description

Two different test sets and, hence, trainings, were designed to validate the software version of the system. The first one, with a larger set of participants, the training and tests described in section 6.6.1 were used as training for the real system. In the second one, four participants agreed to become experts by means of longer trainings and tests, following the steps described in A.III.3.1.

The tests completed by the participants and/or experts are the following:

- *Test 1: Table test.* The test done by all the participants consisted on identifying the position of different objects on a table, without touching or seeing. The setup and the combinations tested are described in A.III.3.1.1. The experts completed this test 4 times (the one at the beginning, they are computed as regular participants for statistical analysis).
- *Repeated tests of the experts.* The experts repeated the Table test in three more occasions, as check points to measure the effect of the training.
- *Test 2: Unknown room.* The experts where left, blindfolded, in a room with unknown configuration of soft obstacles, to move freely trying not to crash against these obstacles, and reporting whenever they find something.

- *Test 3: Pose estimation.* The experts were asked to guess the pose of the experimenter among three possible cases: standing up, sitting or kneeling in front of them. They had to do it 3 times for each pose, randomly distributed (but in fixed order for all the experts), hence, 9 times in total.

To these tests, we have to add the survey completed by both regular and expert participants, as well as the focus group created with the experts.

7.1.3.2. Results

The presentation of the results will follow the structure of the experiments exposed in the previous section.

Test 1: Table Test

The real environment experiment consisted, as said, in 9 concatenated tests locating two or three objects in a 3×3 grid on a real table at 20 cm from the closer edge of the table after a short training to get the reference of the pointing directions of the cameras. 27 people participated of this experiment (one failed test was not considered). The participant had to locate in rows and columns the objects. Each time they located an object away from its real position, two errors were marked (one because of the false positive –locate an object were it is not- and another one because of the false negative –no localization of an object were it indeed was-).

The average number of errors was 33.69% (ranged between 18.05 and 51.39) with a standard deviation (s.d.) of 7.58%.

Marginally significant Pearson correlation was found between the time required to complete the test and the visual impairment (.345, p=.078) although the one way ANOVA mean comparison was not significant (F(2,24)=1.952, p=.164). This relation is shown in figure 7.2.

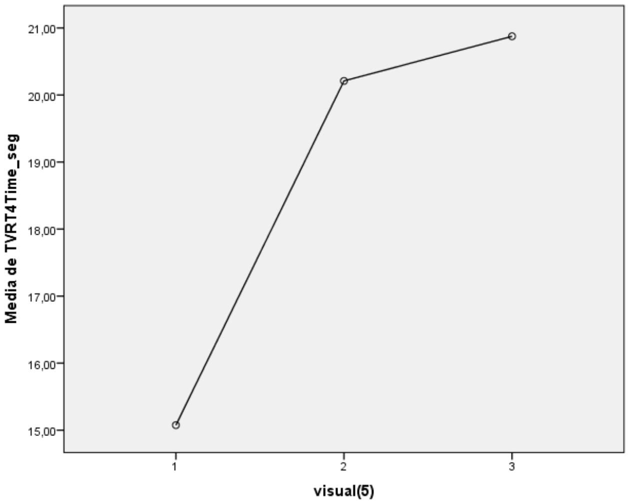


Fig. 7.2. Time to complete the RE test against the visual impairment.

As expected, it seems to be a relation between the use of computer and the number of errors, as shown in figure 7.3.

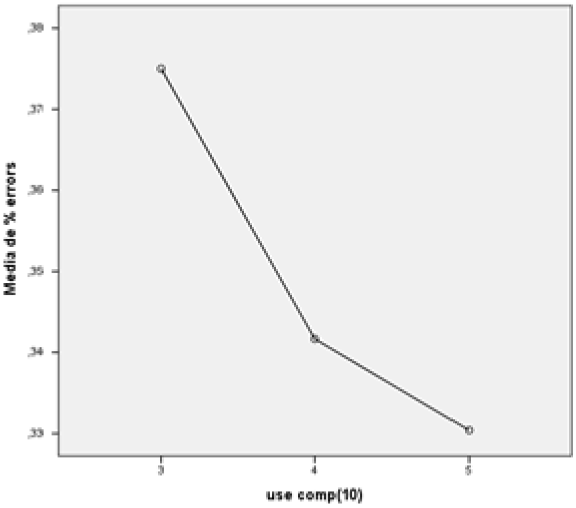


Fig. 7.3. Errors in the RE test against the use of computer. (For this figure, the only sample with use of computer level of 2 has been eliminated).

No statistically significant differences were found when comparing the means of the errors and the profile level. However, as this is an important result to be discussed, this relation is shown in figure 7.4.

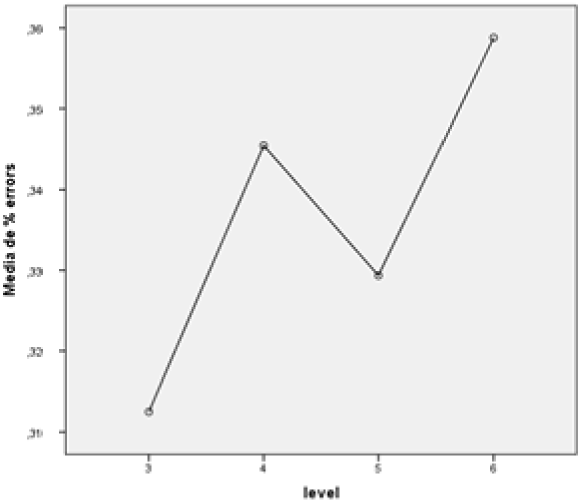


Fig. 7.4. Errors in the RE test against the profile level.

Attending at each cell, the share to the total errors is not equally distributed all around the table. Figure 7.5 and table 3 show the % of errors due to each cell

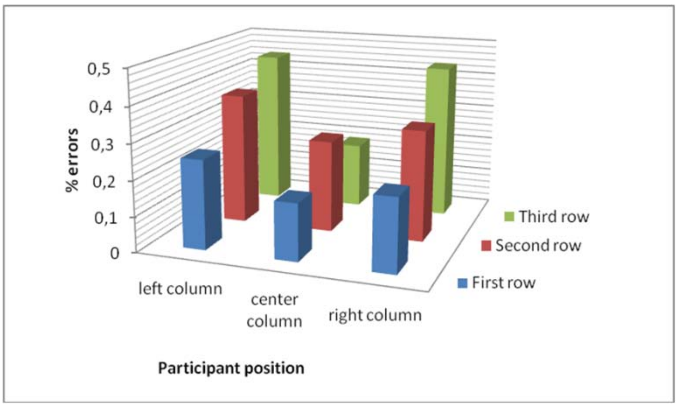


Fig. 7.5. Errors in the RE test, spatially distributed in the table grid.

	left column	center	right column
third row	.44	.18	.44
second row	.37	.26	.31
first row	.25	.16	.21

Table 7.1. Average errors in each cell.

Repeated Tests of the Experts

As previously explained, the experts repeated the RE test in three occasions, as check points to measure the improvement of their skills in artificial vision. Figure 7.6 shows the progression for each one of them (proportion of errors in absolute scale)

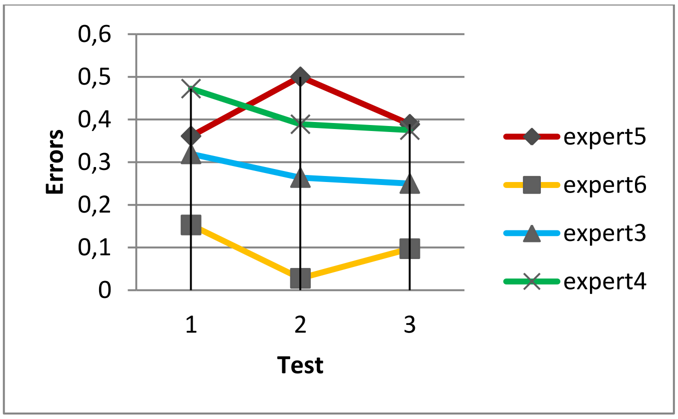


Fig. 7.6. Progression of errors in the RE table in three different moments for each "expertX" ("X" indicates the profile level of each expert).

Test 2: Unknown Room

Figure 7.7 summarizes the correct detections, miss-detections, false positives and "STOPs" (not included in miss-detections) needed to avoid crashes during the experiment of 20 minutes of blindfolded free walk around the room.

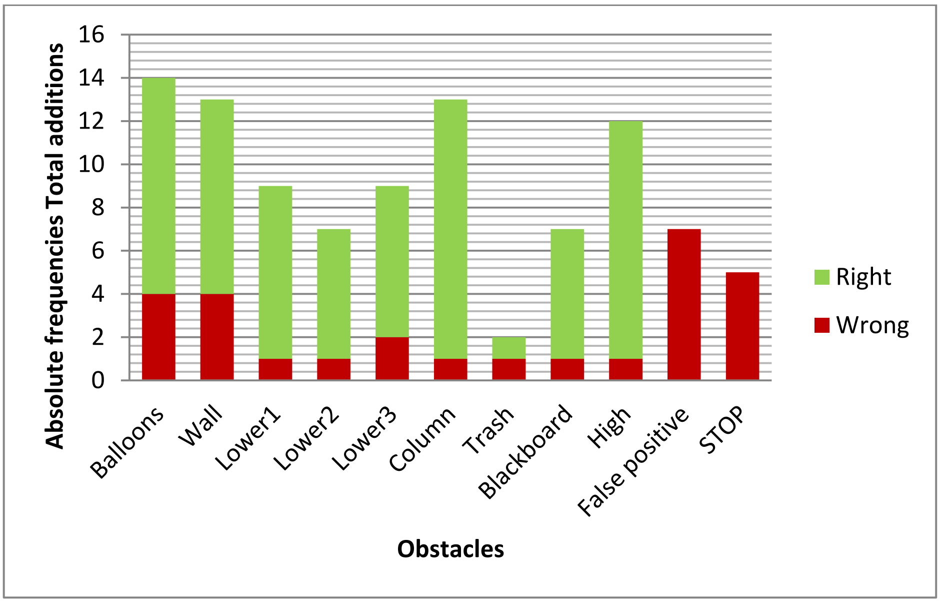


Fig. 7.7. Total number of right detections (in green) and miss-detections (in red) for each obstacle. False positives and STOPs are shown also in red.

The number of errors (miss-detections, false positives and STOPs needed, all Wrong detections) and correct detections (Right) for each profile is shown in table 7.2.

	Profile 3	Profile 4	Profile 5	Profile 6
Right	11	17	17	25
Wrong	6	10	10	5
% errors	35.3	37.0	37.0	16.7

Table 7.2. Average errors in each cell.

Test 3: Pose Estimation

Nine random (but keeping the order in each test) poses were performed by the same experimenter in front of the experts, blindfolded and at 1 m distance, 3 times each pose. Figure 7.8 summarizes the correct and erroneous estimations of the poses.

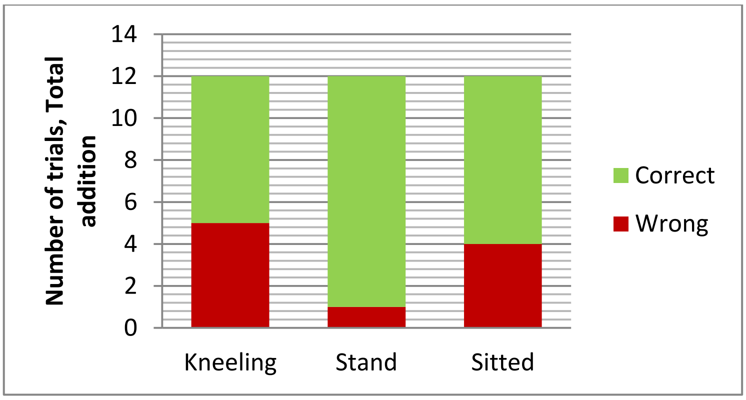


Fig. 7.8. Addition of correct pose estimations (in green) and incorrect estimations (in red) for all experts.

Participants Survey

Participants fulfilled a survey to evaluate subjectively the experiences, answering to questions 37 to 42 of the survey shown in A.II.3.4 or through this link. These questions versed about ease, usability, problems and limits of the system, as they were perceived by the participants.

Significant correlations were found between the profile level assigned to the participant and the perception of the speed of the system (Pearson's correlation index of -.396, p=.037), the thoughts put to understand the scene and the visual impairment (Pearson's correlation index of -.452, p=.016) and the effects of the distortions of the real system and the perception of the speed (Pearson's correlation index of .489, p=.008), the effort needed (Pearson's correlation index of -.485, p=.009) and the number of objects from which it was hard to be understood (Pearson's correlation index of .482, p=.009). Finally, the though put into the experiment and the perception of the speed are also highly correlated (Pearson's correlation index of -.557, p=.002), as well as the effort needed and the number of objects easily detected (Pearson's correlation index of -.587, p=.001).

A marginally significant comparison of means was found when comparing the perceived speed of the system and the level assigned (ANOVA comparison, F(3,24)=2.338, p=.099). Figure 7.9 shows this relation.

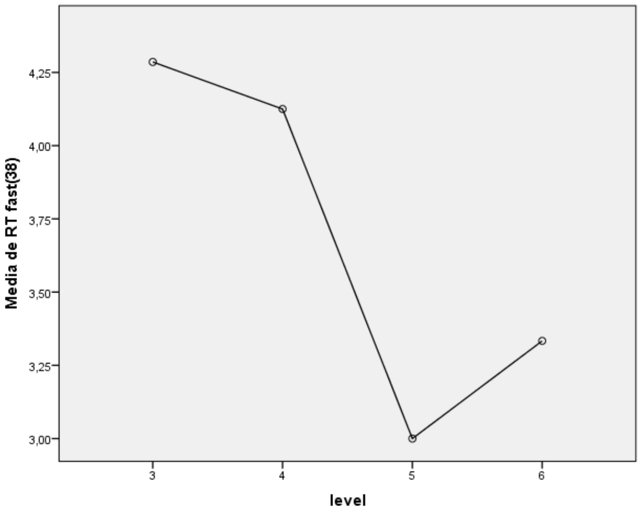


Fig. 7.9. Perceived speed against the profile level assigned.

The comparison between the thought put and the visual impairment was also marginally significant (F(2,25)=3.345, p=.052), and it is shown in figure 7.10.

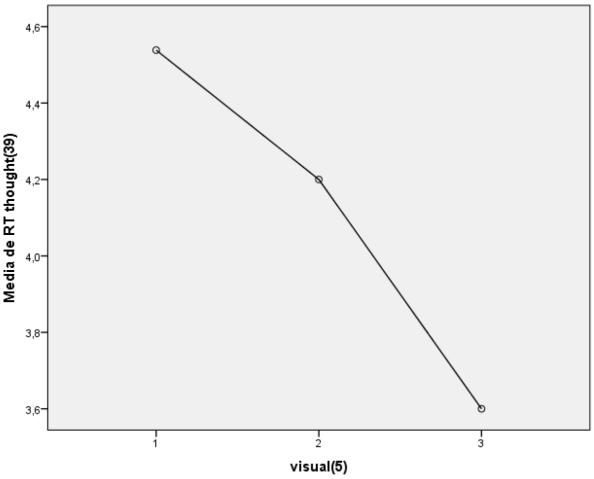


Fig. 7.10. Though put in the experiment against the visual impairment.

Other significant mean comparisons were between the perception of the effects of the distortion of the real system and the perception of its speed (ANOVA F(4,23)=4.214, p=.011) and figure 7.11, the effort needed (F(4,23)=3.52, p=.022) and the number of objects (F(4,23)=3.215, p=.031).

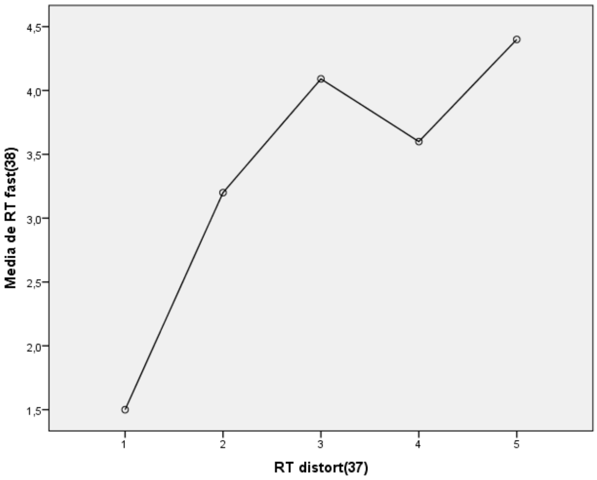


Fig. 7.11. Perception of the speed against the perceived effects of the distortion (in this axis, 1 means high negative effects and 5 no effects).

Experts Survey

The experts were asked, as well, for their evaluation up to now, of the SW version system after the training. The mean of the answers, in a Likert scale (ranged 1-5) are shown in the following table:

Question	Mean
The whole training helps for a vision system	4.25
The whole training helps for a mobility system	4.25
Moving in the room was not stressing	3.75
I felt confortable	4
It is easy to estimate the height of the persons/objects	3
Lower obstacles are still hard to be perceived	4.25
Moving persons are confussing	3
I arrive to "see" the objects	4.75
The errors of the system generate me fear	2.75
Level of trusting on sounds perceived (1: no trusting at all, 5 complete trust)	3.75
My opinion about the system has improved after becoming expert	4.5
I would feel safe using this system in open and unknown spaces	3.25
The system works as "artificial vision"	4.25
The system works as a mobility help	4

Table 7.3. Questions and mean answers of the experts.

Focus Group

A one hour discussion between the experimenter and the 4 experts took place to analyze subjective and qualitative aspects of the system and the perception of the whole process.

Lower obstacles were found to be more difficult to detect. Discussing about the differences between the VR and the RE setups, the RE presented specific problems:

- The position of the eyes (under the helmet and, thus, the cameras) doesn't match with that of the cameras, so the head (the only physical reference of the participant) is not pointing where the cameras are. This displacement causes some errors and difficulties in the RE test, and the position of objects are confused because of this effect.

- The real system has some errors (some of them due to the auto-exposure functions of the cameras that put into troubles the stereovision algorithm) and it is more confusing than the VR system. This problem is critical when pointing to non-textured surfaces, where the stereovision algorithm encounters more problems to be effective (see fig. 5.56). However, it was quite easy for them to know if there was something there, one of them said.

Two main strategies to know what is in front of them were discussed: the so called "scanning" (when the participant just focus on one single tone (that of the middle in general) and tries to find the objects with this single sound, and the "holistic", where the user tries to figure out the whole scene attending to the complete combination of sounds. The first one was used by all of them in the table test, they marked. Another one said that s/he used the holistic approach first, and then started to scan the table for more accurate perceptions.

When discussing about the limits of the system, they also found difficult to know what was below their knees when standing up. In this line, the pose estimation test was easier, some of them said, however they agreed that distinguishing between kneeling and sitting was not that easy. One of them said this test was hard. The edges of the table presented important problems and it was hard, they marked, to know whether there was an object or it was empty in these cells. Likewise, one of them pointed out the confusing effect of the vibrato when many objects are in the scene.

The main complain about the profile level was done for level 6: too many sounds and sometimes hard to manage them mentally. The vibrato effect to detect laterality of objects distorted sometimes the interpretation when too many sounds were present, as one expert pointed out. In this way, another one suggested to cut some information before sonifying it, since there was irrelevant noise not usable for mobility but confusing in the whole understanding of the perceived sounds.

They talked and agreed about the utility of walking in the room with the eyes opened with the system as training for the final test blindfolded. They suggested that blind people could use known places to train the system, as well as verbally describing them how their surroundings are.

It was a generalized opinion that the confidence on the system increased with the use, but some of them pointed out the necessity of more practice and the use in everyday tasks, where it can be very useful, as marked by one of them.

As general evaluations, “the mapping (sonification protocol) is fine” one of them said. However, the noise and the mismatch of the cameras and the eyes remain being the most important problem.

7.1.4. Discussion

Some other interesting results were found in the table tests (33.69%, miss-detections and false positives included). It can be concluded, with some restrictions, that the sonification protocol proposed describes, with a short training, the real world. The restrictions are related with the design of the sonification, the hearing limits in the perception and some demographic parameters. These errors are low enough to guess that the system allows providing descriptions of the surroundings quite accurately. The errors were found to be correlated with the use of the computer in the daily life (figure 7.3), which was a hypothesis coherent with the results obtained in chapter 6, evaluating the sonification itself. More interesting is the distribution of the errors against the profile level (figure 7.4). We find in this case an estrange shape of the curve, with low rate of errors for levels 3 and 5, and higher for 4 and 6. Level 5 was already found to be very efficient when the system was used to differentiate the heights (in this test also the position of objects in the perspective of the user), but a decreasing curve was expected to be found. However, level 4 obtained worst results than level 3. The same happened for level 6. This will be discussed later in this section.

We can explain more easily the spatial distribution of errors in the table all around the experiments (figure 7.5 and table 7.1). The edges of the table and lateral walls of the testing room created an important part of the total errors. Likewise, the distance also contributed to the misperception of objects. The lower error rate was found, then, in the first row, center column.

The experts allowed us to reach new limits in the usability and efficiency of the system from many points of view, since larger trainings permitted to them to get more familiar with the sonification and the specific problems of the real system and testing them (see this video as an example).

The unknown room put the experts in the borderline between mobility and artificial vision aspects of the system. The detection rate of lower obstacles (see the "trash", in figure 7.7) or plain surfaces (see "wall" or "balloons" in the same figure) produced most of the miss-detections. In contrast, columns and paper lower or high obstacles were easily detected by all of them every time they faced them.

The pose estimation test (figure 7.8) shows some of the potential applications of the system. Blind people use to see as problematic to find free seats, in bars or restaurants [2]. We found an accurate perception of the pose of a person in front of the participant. Actually, the errors, with one single exception over 36 tests, were due to the confusion between kneeling and sitting (which is not that relevant when looking for a free seat).

Another hypothesis could be barely tested, due to the small amount of experts performing a longer training, and it can only be qualitatively evaluated. In figure 7.6, we see different tendencies of the different experts. Three of them increased their performance from the first to the last test and two of them had some inverse tendency at some point (increasing the number of errors in some consecutive tests). We should point out that 5 or 6 hours of training for an artificial vision is still a very short one, given that the reconfiguration of the brain exploiting the cross-modal plasticity needs much more time to appear (see [3] for more details). This progressive test should be repeated in a larger sample; however it makes sense that there is a negative relation between the hours of use of the system and the errors made.

When asking the participants to subjectively evaluate the RE experiment, they offered important results, many of them statistically significant. Figure 7.9 shows the evolution of the perceived speed and the level, finding the system faster in the lower levels down to 5, with the change of the progression for level 6. This can be explained as follows: lower levels demand from the participants a lower effort, so the system is perceived as faster (they understand faster actually). For level 6 the tendency is inverted, maybe because users in this level didn't even arrive to understand completely the sounds and, thus, perceived the speed of the system more objectively. High correlations were found between the thought put in the experiment and the perception of the speed. On the other one, important correlations (with p-values below .01) were found when comparing the effects of the distortions of the real system and the effort needed (positively related; the correlation was -.485, but the Likert scale of the evaluation of the distortion was build inversely: 1- distortions are not disturbing at all, 5-

distortions are disturbing). The distortion affected in the same way to the number of objects the participants thought they arrived to identify. Finally, high correlation (significant at a level of 99%) was found between the effort needed and the number of objects. Another interesting result, coherent with the results shown in chapter 6 is shown in figure 7.10. There, we can see how the visual impairment doesn't help to the usability of the system (let's remember that this is not a final conclusion because the demographic properties don't allow us splitting the effects of the different variables that interact in these kind of systems). A similar tendency can be found when comparing the time to perform the RE test and this factor. The lower is the vision, the longer it takes to guess the combination of objects. We could relate this with the effort needed to understand the scene (as also shown in figure 6.96).

When asking the experts to evaluate the training, they agree that longer use of the system make them feel safer, and the usability increases, as well as their opinion about the system. However, they still detect some problems, as tiny and lower objects, and the estimation of the height of objects and persons. Anyhow, they arrived to "see" the objects (a 4 in a 5 levels Likert scale).

The focus group allow reaching deeper levels of understanding of the experience of the experts, with a qualitative approach to improve the system in further implementations: The main problems were found in the complexity of profile level 6, and in the noise of the real system to produce reliable sonifications. The suggestion of cutting some information (the less relevant one) to increase the usability should be taken into account in further implementations. This is related with the result shown in figure 7.4: level 6 seems to work in another way than the other 3 ones. This may be fixed with a longer training that allows the users getting familiarized with the complexity of this level.

Somehow related with the second aspect, the noise of the real system, the mismatch of the directions of eyes and cameras seems to be a critical problem, which should be solved in the next prototype.

Experts found two strategies to pass the tests. The scanning strategy, although provides less information, produces a more accurate one, at least in the first moments of use of the new product. This approach should be used, in the future, with lower profile levels, to avoid irrelevant information. However, and once again, we think that longer trainings will help to develop an intuitive holistic use of the system. This hypothesis needs to be checked in the future. The system, still, has its own errors, although they are not that relevant (the answer to the question "The errors of the system generate me fear" gave a mean of 2.75, and a mean of 3.21 when the participants were asked to evaluate the effect of the distortions of the real system (being 5 no effect at all, as said).

We can affirm, with all these data, that hypothesis H3 has been widely confirmed in these results.

The main limitations of this study are the same than those exposed in section 6.7, so they will not be repeated.

7.2 Hardware Implementation

Computers are general purpose hardware. Likewise, for a portable device, we must reduce the power and weight requirements of a computer, even of a laptop. Moreover, as stated in [4], hardware-based image processors achieve much higher rates than software versions running on standard PCs. We can find many works implementing hardware solutions for stereovision algorithms, such as [5-8]. However, this is not always the case. Depending on the power and the technology used, the processing rate may be equal or even lower than the software version in a PC. On the other hand, the price may be, also, lower.

For these reasons, we will analyze the different options in the hardware, propose a design and implement a physical device with which will perform the final user tests.

7.2.1. Limitations allowed to the Hardware Version

The main goal of the system is to help the visually impaired to avoid accidents when they are travelling in unknown environments. With the help of the user tests made with the software version, we can also reduce some other requirements. Finally, we should remember that the low cost was also an important constraint. Thus, we can summarize the final limitations to the following list:

- Level 6 will not be implemented. Given that it didn’t show an important improvement in the perception of obstacles, but increased the feeling of annoy, it was not considered in the hardware version.
- Images of 160×120 pixels still present the most relevant characteristics, reducing the computational load required.
- The consumption power must be low, to achieve at least a couple of hours of autonomy with a relatively light battery.
- Not every hardware system is able to produce MIDI sounds. We will keep some basic features of the proposed sonification as mandatory: 8 discrete panning positions, with different low frequency modulations (tremolo effect), up to 8 different pitches (for level 5), correspondence between distance and loudness and correspondence between distance and timbre.
- Not needed to represent the 64 sounds in the level 5. Discrimination of danger (distance) and sonification of the closer ones.
- Total cost under 500€.

7.2.2. Technology Study

In the electronic manufacturers market we can find thousands of devices and microchips with many different performances, prices, encapsulations, etc. Among them, we will compare some devices for the image sensors, as well as four technologies for the image processing hardware implementations.

Mobile and low cost technology has experimented a huge grow the last ten years, due to the democratization of mobile phones and other multimedia technology. Thus, we can find many devices whose characteristics regarding power consumption, speed and, over all, price, would be impossible just a few years ago. The main advantage of every new product is that it can use new technology at reduced costs.

7.2.2.1. Available Hardware

Five technologies and architectures evaluated are the general purpose microcontrollers, the Field Programmable Gate Arrays (FPGA), the Digital Signal Processors (DPS), smartphones and micro-computers.

Microcontroller

A microcontroller is a whole system on a chip (SoC), which presents a microprocessor, RAM memory, storable memory and input/output (I/O) ports such as digital ones, analogs, serial communication ports, among others.

This technology presents some problems:

- Serial computation: every instruction is executed sequentially, and each one of them has to be loaded in the CPU to be executed. This architecture limits the processing speed with which uC can implement certain tasks.
- Shared hardware resources: Memory addresses, internal buses, the ALU, registers and so on are limited and shared resources available for the program running on the microcontroller. This fact limits the performance of the whole system.
- General purpose ALU: the arithmetical units of uC use to be designed for general functions, requiring some of them, as divisions, high computational resources, with the loss in performances achievable by this technology. This constraint is not always present, as we will see.

However, some advantages should be cited about this technology:

- Easy programming language (usually C).
- Low-cost solutions, finding uC from < 3 euro.
- Low-power consumption. There are microcontrollers needing around 3mW (Cypress PSoC3/5 [9]) but with limited performances.
- Reprogrammable on the final device, allowing program changes without extracting the IC.

One test over the MCU Toolstick F330 DC [10] has been performed, to evaluate the hardware performances of uCs. The chosen uC has a 25MHz 8051 core and 256 bytes of RAM. The uC was programmed with a simplified version of the program, which processed two lines of 20 bytes and computed one 2.5D line of 20 bytes. This process was repeated 16 times (to process 360 bytes per line) and 240 times (to simulate the processing of 320×240 bytes images). The rate of this whole processing was around 1 fps. Given that the system must synthesize many synchronic sounds, and control the cameras, the final rate will be much lower.

FPGA

The FPGAs represent a change in the paradigm of hardware transposition of PC programs. On the one hand, they are not devices for software programming (as the uC are), but hardware programming. Differences are relevant. When a FPGA is programmed, its physical structure changes, so what we are doing is creating paths and physical signals, instead of only variables, as it is the case of the software programming. Another consequence of this architecture is the possibility of parallel processing. Several signals can be processed in parallel in the same FPGA.

With the miniaturization revolution of the latest years, and the democratization of the 45nm technology, FPGAs can achieve millions of gates, DSP internal blocks and other features further than the logical connection of gates. FGPAs need to be programmed in some hardware description language, as the VHDL, among others. This is a drawback, since it is mandatory to translate, if finally we choose this technology, our program to another language. However, there are C-to-VHDL compilers and other translation software, so this task may not be problematic. A final drawback of this option is the relatively complexity to produce and weld, and the low-level synthesis of sounds.

DSP

The DSPs are the special purpose uC, designed and optimized specially for signal processing, such as images, sounds or other forms of information. These devices reproduce some of the uC limitations. However, some of them are solved or minimized, by means of some special purpose hardware, such as multipliers/dividers, which can perform these tasks in one single clock event. Actually, they are usually designed for specific tasks [11].

The main problem is the relatively complex package and managing board. For example, the TMS320VC5506 [12], a general purpose DSP, with 128KB of RAM (one of the biggest), has a 100LGFP package, which means that 100 pins must be welded and configured.

Smartphones

An interesting option revised was the use of smartphones, specifically those with two cameras and powerful cores. We purchased a HTC EVO 3D smartphone, based on Android 2.3 [13], with 1.2GHz core and 1GB of RAM, which should be enough for our purposes. This option presented, however, some important problems:

- First of all, the price. This device (and those similar, such as the LG OPTIMUS 3D) costs around 450€.
- We found that the two cameras were not exactly the same, and the autofocus of one of them is completely out of control and, hence, the stereovision algorithm couldn’t achieve fair enough results.

- The MIDI synthesis in real-time is impossible under Android. Other options were to use PureData (which will be explained later).
- The image processing was not that fast, mainly due to the several concurrent processes running in a smartphone, so it is not dedicated to the main tasks. The frame rate found was around 3fps with no sound synthesis.
- Translation needed from C to Java.

Micro-computers

In the year 2012, many new micro-computers became commercial, with some relevant characteristics, being the main one the flexibility and the low cost. Many of them run under Android (such as the MK80 [14], the RK3066 [15], the Z902 [16] or the A02 [17]) with costs between 40 and 120$. We found also another one running on Debian, the Raspberry Pi (RPi) [18], much more flexible and cheap. These devices use to have drivers to manage USB cams, and analog outputs to directly send the synthesized sounds to a jack stereo cable. In the case of the RPi, the programming is done over C and, thus, no translation is needed.

7.2.2.2. Comparison

The decision shall be taken regarding some constraints:

- Easy to program.
- Easy to integrate in the layout.
- Low-cost.
- Low-power.
- Fast enough for the required processing.
- Enough RAM memory.

The following table shows the main characteristics of some average devices of each technology.

	uC	DSP	FPGA	Smartphones	Micro-computer
Example	MSP430	TMS320F2809	XC3S1400AN	HTC EVO 3D	RPi
Speed	4KHz-8MHz	108MHz	66MHz	1.2GHz	1GHz
RAM Memory	128-2048B	64 KB	176 KB	1 GB	512 MB
Prog. Language	C	C?	VHDL	C/PureData	C/PureData
Packing	14-48 pins	LQFP100	TQG144	-	-
Consumption	7mW	160mW	1W	.74W*	3.5W
Price	6€	14.82€	14.94€	450€	30€
Complexity**	4	7	8	3	1
Source	[19]	[12]	[20]	[13]	[18]

*Time of conversation: 7h 45 min with a battery of 1730 mAh. This represents around .74W. This doesn't take into account the extra load due to the image and sound processing.

**Complexity is a subjective measure about the ease of implementing the final prototype with each technology, ranged from 1 (very easy) to 10 (very difficult).

Table 7.4. Hardware technologies comparison.

The final hardware device chosen for the project is, finally, the Raspberry Pi, because of cost and simplicity reasons.

7.2.3. Implementation

The RPi is a system-on-board with full computing capabilities. The main features of this board are the following [21]:

- Broadcom BCM2835 (CPU + GPU), with an ARM1176JZF-S core running up to 1GHz (with overclocking. Standar speed: 700MHz).
- 512MB of SDRAM.
- 2 fully operational USB 2.0 ports.
- RCA Composite video and HDMI port.
- 3.5mm jack output analog audio.
- 8GB SD-card of storage.
- Wheezy Debian operating system.
- 10/100 wired Ethernet RJ45.
- GPIO pins, serial, UART and other peripherics.
- Micro-USB power source.

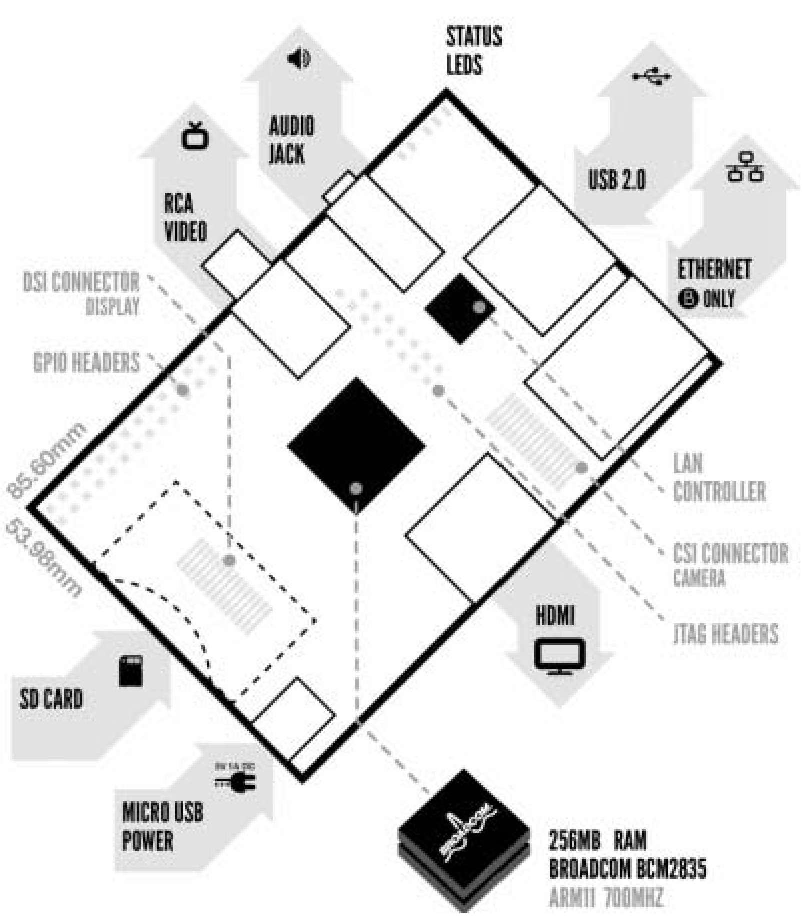


Fig. 7.12. RPi connections [21]

7.2.3.1. Software Design

The Wheezy OS based on Debian allows the compilation and execution of ANSI C programs. An ANSI C program, using the "imgproc" library [22] to capture color images, creates the depth map over 160×120 grayscale images, simplifies it to a 8×8 2.5D image and generates messages following the next diagram:

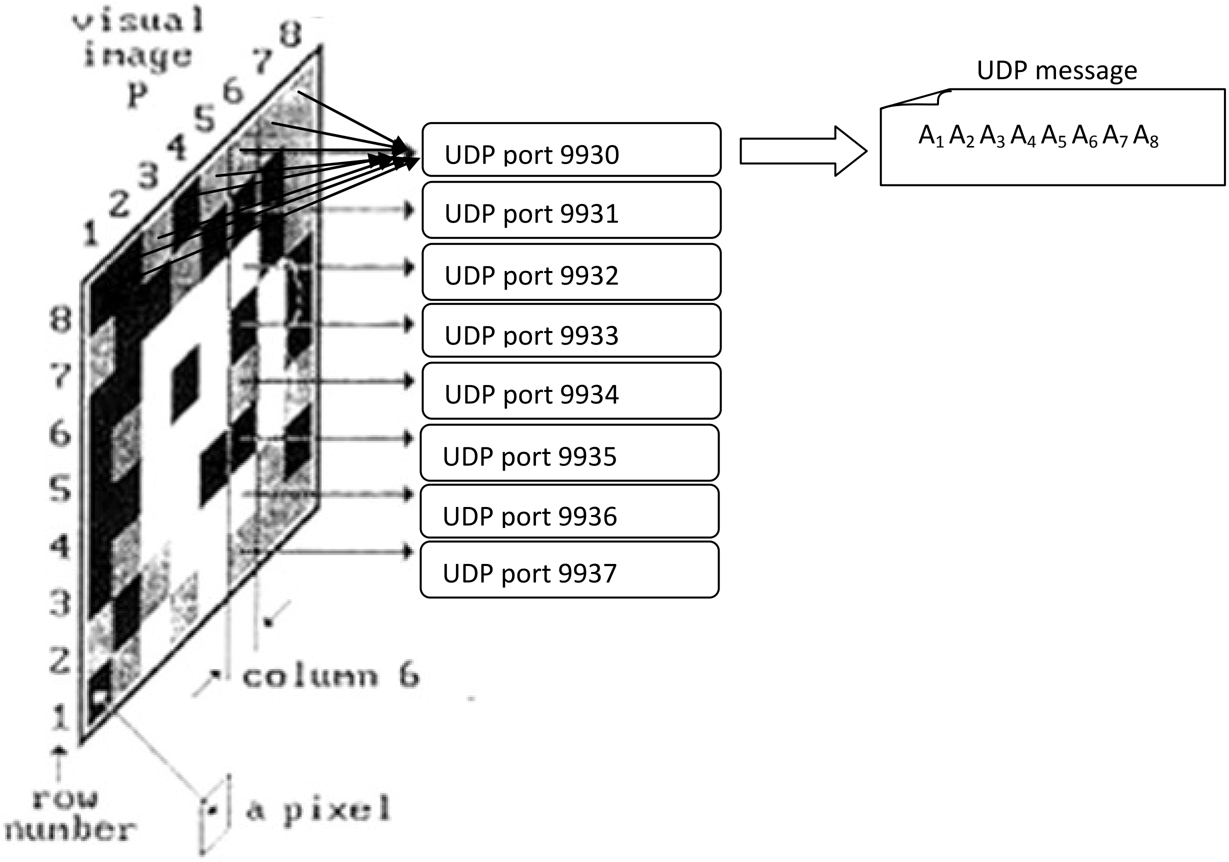


Fig. 7.13. Messages formation following a row structure.

In the fig. 7.13, A_i represents a char number (8 bits) with the gray scale of the *i-th* column. Each row produce a set of A_i values, grouped in a single UDP message. These messages are sent every frame is produced and analyzed, to 8 different ports of the localhost.

Another program, implemented in PureData [23], receives these messages, using the following patch, named *testUDP5.pd*:

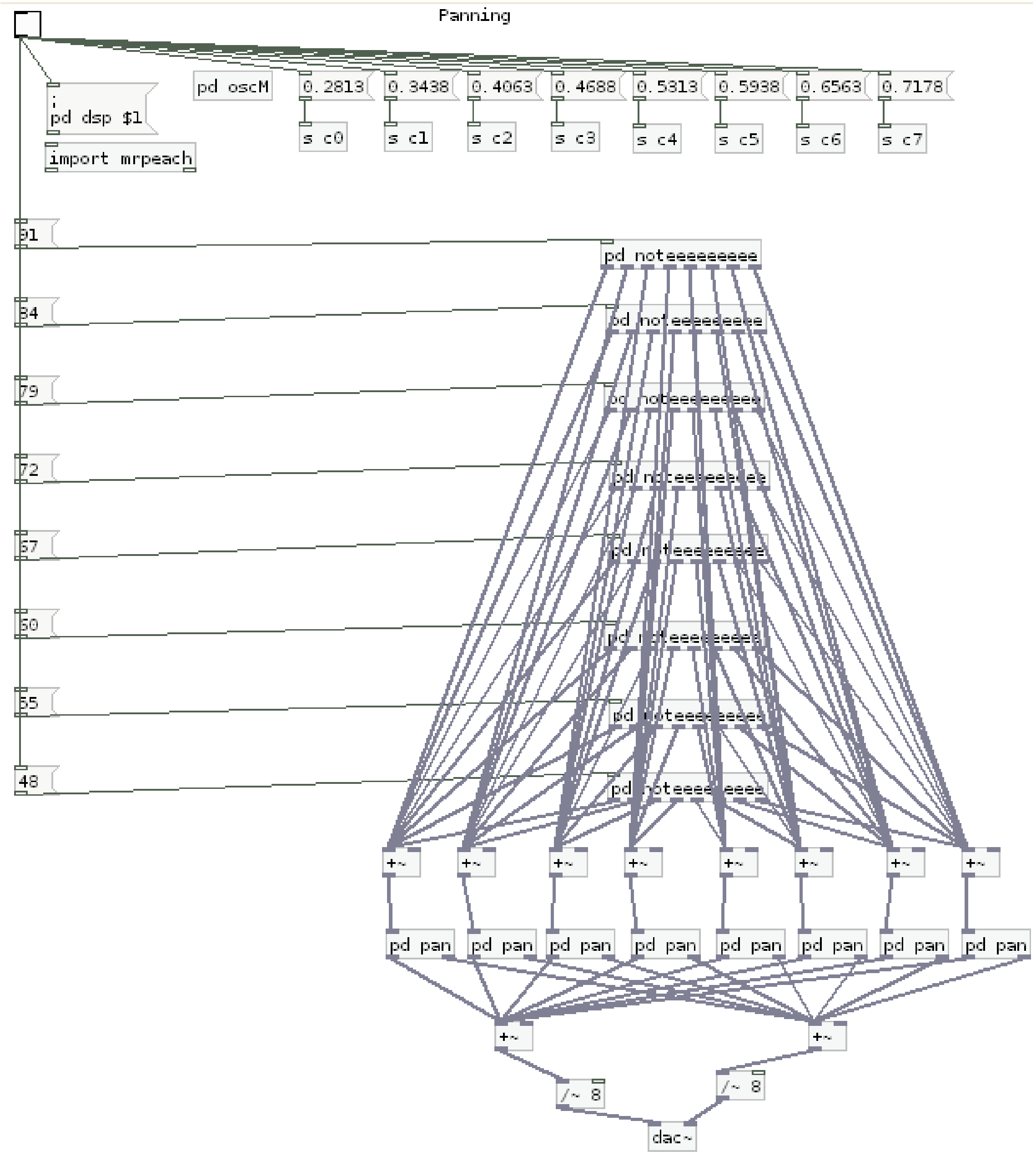


Fig. 7. 14. Puredata basic patch, implementing a receiver for each row and establishing the values for the notes (left column of values) and the panning (horizontal row of values).

Each receiver, called *noteeeeeeeee* (the number of “eeee” is just to force the program to show all their outputs) is designed as follows:

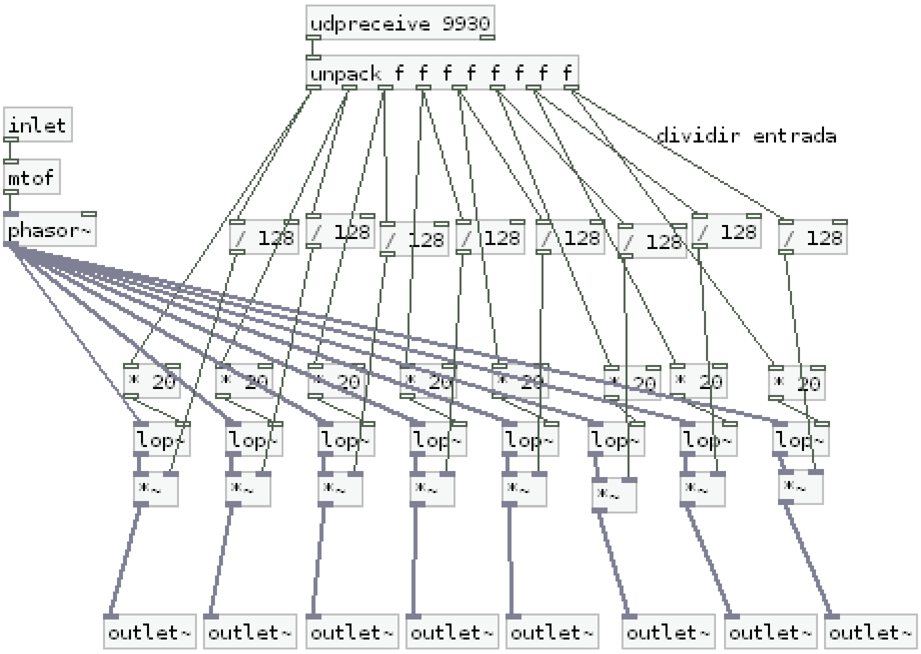


Fig. 7.15. noteeeeeeeee subpatch.

The *noteeeeeeeeee* subpatch receives a number representing the correct frequency (i.e. each *noteeeeeeeee* is different, depending on the row each one implements) and an UDP message from the correct port, in the case of that one shown in figure 7.15, from the 9930.

The number produces a wideband noise (*phasor~* object). The UDP message is unpacked, the value divided by 128. The amplitude received is used to modulate the cutoff frequency of a low pass filter that transforms the noise. Each column produces a final noise, which is sent to the following stage.

The *pan* subpatch is implemented as shown in figure 7.16.

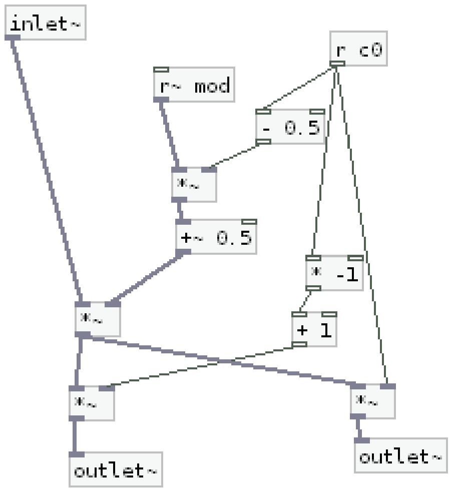


Fig. 7.16. pan subpatch to produce panning effect.

This subpatch, again, is different for each column. The shown one is that of the leftist position, represented by the variable “c0” read at the beginning.

The puredata is automatically initialized on boot with the expression:

```
sudo nice -n -20 pd-extended -rt -nogui -alsa sonif.pd
```

In Debian, this allows executing it with the maximum priority and in real-time mode. Otherwise, the DAC of the RPi produces permanent flicker noise.

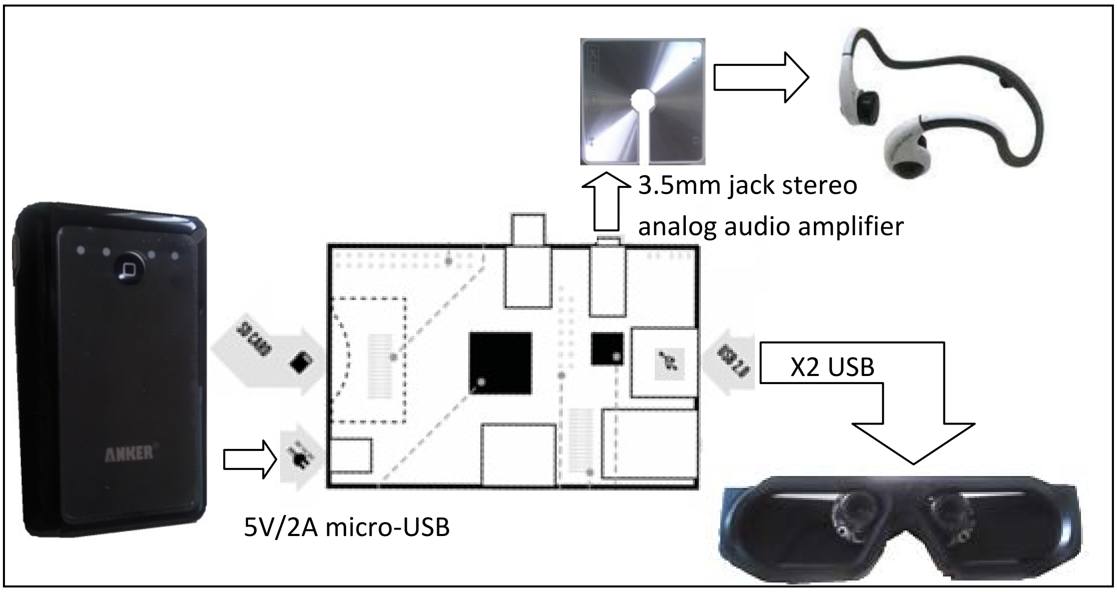
7.2.3.1. Hardware Connections

The webcams of the hardware version were those used in the software version [1].

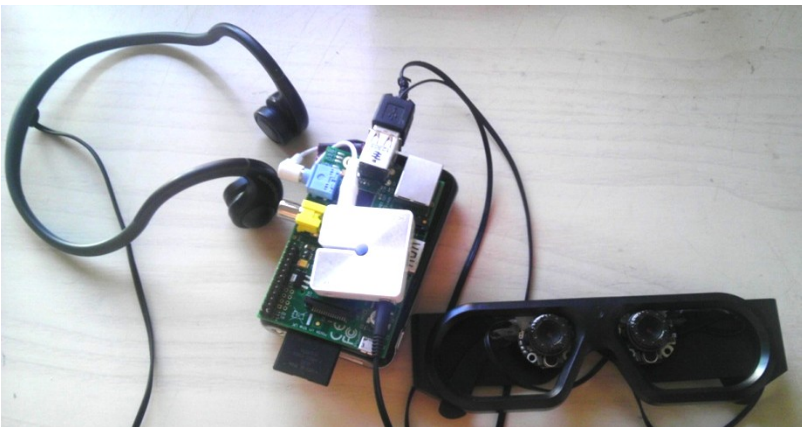
The bone conduction headphones used were the AudioBone 1.0 [24], with a frequency range of 50Hz-12kHz. The bone conduction headphones don’t provide a sound loud enough to be perceived in noisy environments. Thus, an amplifier is needed to give power enough to the sonification. Thus, an amplifier is needed to provide sounds with the proper loudness. The Audio Amplifier/Splitter Jack 3.5mm [25] was chosen, although not the cheapest one, it was the easiest and fastest to be obtained, and it has its own battery, with an autonomy of 12 hours.

The battery used in the final device is the Anker Astro2 Battery [26], with a capacity of 8400mAh and a limit of 2A in the output, which is far higher than the RPi maximum input current. This should give us an autonomy, again, of 12 hours without any recharge.

The final scheme is represented in the following figure.



(a)



(b)

Fig. 7.17. (a) Final interconnection of the different devices and (b) real aspect of the first prototype of the ATAD.

7.2.4. Evaluation

Two main evaluations were carried out. The first one, a quantitative analysis, focused in the prototype characteristics in several fields: performance, consumption and cost. The cost is also evaluated for the whole project.

The second one was a users' test, done with the help of 8 blind participants, who used the device in real environments and gave their opinion about how it works.

7.2.4.1. Quantitative Evaluation

The quantitative evaluation deals with intrinsic characteristics of the system and the project itself. It will be divided in three categories: time, consumption and cost.

There is no way, at this point, to objectively evaluate the errors of the algorithm, due to the lack of the depth truth

Time Analysis

Time performance is important in any device, especially those supposed to work in real time.

The RPi, fully running (that means, with the puredata and the image processing working in parallel, full load) sonifies at around 10.1fps (mean over 4s working) and 99ms. This time can be divided in that used by the sonification and that used by the image processing. Running the device without the puredata thread, the image processing alone (with the other linux kernel threads) run at 15.5fps or 64.3ms per frame. We can derive the speed of the sonification in 34.7ms per frame or 28.8fps. It takes, finally, 73s to boot.

Consumption Analysis

As specified in section 7.2.3, the battery used during the tests had a theoretical capacity of 8400mAh. This battery takes 7.5h@1.2A to get charged.

This load gives an autonomy of 10h, what represents a consumption of 4.2W, slightly higher than the producer datasheet.

Cost Analysis

The hardware prototype cost at consumer end prices and its distribution is shown in the following figure.

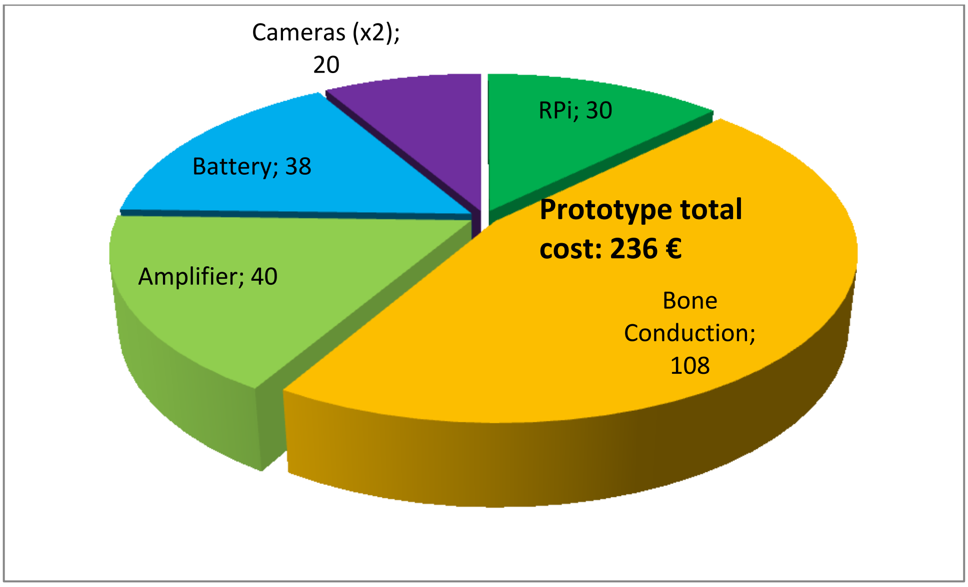


Figure 7.18. Hardware cost.

No assembly is needed further than connecting jacks and USB cables.

However, the main cost of the project is due to the research effort during 3 years and a half, stays and other research expenses:

Concept	Cost
Grant (6630 hours)	54,000 €
Two stays abroad of 3 months each	6,680 €
Conferences attendants (twice)	4,200 €
Devices and laptop	800 €
Total	65,680 €

Table 7.5. Human and research costs.

7.2.4.2. Users Test and Validation

In the tests participated 8 completely blind people (with at least 22 years of blindness), 4 females and 4 males with ages ranged between 22 and 60 (mean 41.38). Seven of them had normal hearing, and one of them used a hearing aid in her right ear. Three used guide dog in their displacements, the rest of them white cane.

The details of the experiments, as well as the surveys used to gather all the relevant information after each one of the 4 tests, and some videos recorded during the tests, are described and accessible in Annex IV.

Test 1: Static Images and Virtual Reality Training

The first test was implemented over static images and a virtual reality environment, to help the participants to get familiar with all the profiles, asking them to choose their favorite one at the end, so they could perform the rest of the tests in this profile.

The average time used in this test was 2 hours 20 minutes (ranged 1h15 to 3h). When asked if their comprehension of the sounds had become better during the session, the average response in a 5 options Likert scale (meaning 1 lower comprehension at the end of the session, 5 a better one) was 4.3 (ranged 3 to 5).

The ease in localizing objects (1 very difficult to 5 very easy) in each one of the user-centered axis, horizontal, vertical and distance, presented a mean value of 4.3 (standard deviation –s.d.- of .5), 4.5 (.9) and 3.7 (1.3) respectively. This last variable produced higher disparities. Moreover, this last variable presented a significant correlation with the level chosen (Pearson's correlation: -.737, bilateral significance: .037). This represents an increase of the difficulty to understand the distance when the level gets higher. Looking for other significances, we found marginal significant results in the correlation of the ease of horizontal and vertical localization with the age: -.626 ($p = .097$) and -.702 ($p = .052$) respectively.

There were also disparities in the opinions about the ease to understand the static images, with a mean of 3.75 and a s.d. of 1.3.

A more consistent opinion was found when asking for the help of the dynamic images to understand the sonification, with a mean answer of 4.75 (.5).

However, more complex scenes such as corridors, doors, etc. were found more difficult (mean of 2.75 and s.d. of 1 after asking for the ease of understanding such scenes).

The comfort level was 4.37 (.7).

When asked for the bone conduction (BC), the response about the ease to perceive the real world's sound was 4.9 (.3), and when asked if they thought it was an acceptable solution for this technical aid, the answer was 4.5 (.7).

Finally, they found the BC not so uncomfortable, answering with a mean of 2 (1.6) when asked if they felt it uncomfortable.

Test 2: Objects on the Table

In this second test, participants are asked to use the real system, without moving, to localize objects in the table and estimate the pose of the experimenter. This test lasted 68 minutes (ranged 45 to 90).

Again, the general opinion is that the comprehension of the sounds increased during the session (4.25, s.d. of .9), although this value was marginally (and inversely) correlated with the age (Pearson's correlation: -.654, $p = .078$).

The same grade of consensus was found evaluating the presence of “noise” (fake sounds that appear without any reason, and false negatives), i.e. 4.37 (.7). However, this is not translated in a decrease of confidence on the system (2.88, s.d. of 1.3) although it makes the sounds slightly harder to be understood (3.75, s.d. of 1.5).

The device is not perceived as slow (2.5, s.d. of 1.2), and participants thought, at this point of the tests, it could be useful for them (3.9, s.d. of 1.1). The weight and form of the system seem OK (4.1, s.d. of 1.1) and the glasses and BC together don’t bother too much (2.5, s.d. of 1.2).

The configuration of objects was not found to be easy to be understood (2.5, s.d. of .9). The difficult to find the smaller ones was slightly over the neutral answer: 3.63 (1.3). Finally, many objects generated confusion among the participants (3.87, s.d. of 1.7). In this point, a negative marginally significant correlation (-.657, p = .077) was found between the educational level and the confusion generated.

Regarding the pose estimation, with four different heights in front of them, it was not found it to be so easy (3.25, s.d. of 1.1).

Test 3: Moving at Home

This test is designed to take advantage of the knowledge of their own houses to help the training with the system.

For around 45 minutes (ranged 35 to 60, s.d. of 7,4 minutes), they were asked to walk and focus on different places, spaces or corners of their houses to get used to the sounds.

The improvement of the comprehension of the sounds during this test was softer than in the precedents. The improvement was felt as being almost neutral (3.63, s.d. of .9), although a marginal inverse correlation was found between this parameter and the age (-.646, p = .084).

The utility of the system in indoor scenarios was not perceived to be so high (2.75, s.d. of 1.5), nor the security supposedly provided by the system outdoors (3.5, s.d. of 1.2). Finally, the utility of the system to help situating themselves in the space was even lower (2.5, s.d. of 1.5). An important result about this parameter is its completely uncorrelation with the educational level (p = 1). The other complete uncorrelation was found between the utility in indoors perceived and the level chosen.

Test 4: Moving Outside

This step tests the real use of the system, at the current development point.

The duration of this test was 47.5 minutes in average (ranged 45 to 60 minutes). In this occasion, the improvement of the understanding of the sounds was higher than in previous tests: 4 (.9), although the system, at this point of development, doesn’t seem to be useful outdoors (2.75, s.d. of 1.7). Marginal negative correlation was found between this and the age (-.627, p = .096).

The confidence on the system, thus, remained the same (3, s.d. of 1.4), and the security feeling decreased (2.9, s.d. of 1). Regarding the first variable, it was found to be marginally correlated with the educational level (.668, p = .07). Regarding the second one, a correlation was found, again, with the educational level (.763, p = .028).

Some different objects or obstacles were found during the test, and the next table shows the ease for detecting each one of them, ordered in terms of ease in detection.

Obstacle	Average ease (s.d.)
Walls	4.13 (1)
Cavities	3.43 (1.4)
Cars	3.38 (1.1)
People	3.25 (1.4)
Awning	3.25 (1)
Chairs/tables	3.2 (1.3)
Trees/semaphores	3 (1)
Mailboxes	2.86 (.9)
Gates	2.86 (.9)
Bars	2.43 (1.5)
Fences	2 (1.1)
Scaffolds	1 (.8)

Table 7.6. Ease of detecting different urban furniture.

They exposed that the system is not useful for mobility goals at this point of development (2.63, s.d. of 1.6).

The subjective assessment about the system remained almost neutral along the process (3.63, s.d. of 1.1), but showed a correlation with the educational level (.713, p = .047). The final assessment after all the tests is slightly higher than the neutral answer (3.38, s.d. of .9).

Other Information Gathered from the Tests

Final questions were posed to the participants so they could give their point of view about the general process.

The real system was not found to be more tiring than the virtual one (1.63, s.d. of 1.2). When asked for the training length, they judged that the training shouldn’t be longer (2.88, s.d. of 1.4). This last parameter was found to be marginally inversely correlated with the gender (-.69, p = .058, which means, following the coding of this variable, that females felt than the training should be longer than the men).

There was a shared opinion about the help of the virtual system to understand the real one (4.62, s.d. of .5), as well as about the BC solution and its capability to allow the reception of real world sounds (4.5, s.d. of .8). Regarding this last question, the opinion was correlated with the chosen level (.75, p = .032).

However, the system did not provide a feeling of security (2.25, s.d. of .9), and a marginal correlation was found between this and the educational level (.64, p = .088).

Finally, every participant stated that she/he would keep using the cane or dog even if they had the system.

Open Questions

The participants had the opportunity to communicate their thoughts about the system after each session.

Gathering the information received, we can build the following table of questions and frequencies (take into account that each participant could expose the same problem in different tests, so the frequency is not *a priori* ranged):

Comment	Frequency
Distance not very well appreciated	13
Fake sounds (false positives)/low accuracy	12
Panning not enough separated in TO	7
Volume control/Automatic Gain Control needed	7
Objects (semaphores, kiosks, etc.) and colors detection and specific functions needed	6
It will work with some improvements	5
Adjustable distance detection	5
Need of an alarm (with specific sounds) when something is close enough	5
(More and) more different sounds needed	4
Vibrations in the middle, softer sounds in the sides	4
Interrupted sounds in the VR	3
Adjustable the field of view sonified and the cameras direction	3
Enough volume in indoors	3
Vibrations not enough appreciated (overall in low pitches)	3
Real sounds (i.e. traffic) impede listening to the system	3
Useful to find darkness/wide empty spaces	3
Complementary to the cane/dog and/or other technologies	3
Better as a pointer, instead of glasses	3
Glasses uncomfortable	3
BC is a good solution to listen to ambient sounds	2
High training/concentration needed	2
Good Price	2
General difficulties with the sounds	2
Uncomfortable to move the head	2
Panning helps	1
Height very well appreciated	1
Increase all the pitches	1
Panning easier in the VR	1
Vibration helps (to the partially deaf participant)	1
It is tired after a while	1
Hard to understand the field of view of the cameras	1
Intermittence of sounds to represent distance	1
Different lateralization in each side	1
Include verbal messages	1
More work needed on the prototype, although good first version	1
Higher speed	1

Good size	1
Skin vibration to help transmitting information	1
Untextured surfaces fail	1

Table 7.7. Open questions after gathering.

7.2.5. Discussion

The real time performance measurement gave a variable measure with an average value of 99ms and 10.1fps, under the initial constraint. Likewise, the sonification, based on puredata, lasted 34.7ms (28.8fps), as shown in section 7.2.4.1, which respects the real-time constraint.

Regarding memory aspects, the HW version of the system could not capture directly the images in grayscale, and they must be stored, temporally, in color bitmaps. Thus, the memory used is increased in 460.8KB, giving a total memory of 538KB.

We also proposed an autonomy for this first prototype of 3h. The final hardware device, with the 8400mAh battery, lasted 10h to consume this energy in fully working mode.

The respect of the hearing system of the user, critical in mobility aspects, was implemented with the already cited BC technology. The reception of this option was very positive (rated as 4.9 over 5 when asked if it was a good solution to compatible the natural sounds and the synthetic ones).

The price was one of the most important constraints, as found in the user requirements interviews (section 3.3.2). A maximum price of 500€ was established. The final cost of the first prototype, with consumer prices, was 236€.

A final merit index was described in section 4, eq. 4.2, with a minimum value of 108e-3. The first prototype obtained a merit index, with the same equation and the obtained results, of 322e-3.

The first subjective result we got, about the ease of understanding the sounds and their correlation with the position, put us face to face against an apparent paradox: the psychoacoustic based parameters (horizontal and distance dimensions) were perceived as harder (4.3 and 3.7 respectively) than the more arbitrary one (the vertical), as easier with an average response of 4.5). We should explain each parameter separately, since they may not share a common reason for their result.

For the case of the horizontal axis, we should think that the panning was designed following the MIDI specification, and correlating the real direction of the object. This worked properly with earphones, as in the software integration. However, in the hardware one, we used bone conduction. Studies are not consistent about the effectiveness of the bone conduction to propagate stereo sounds, providing interaural isolation (see, for example, [27, 28]). We can expect that the perception of the panning was somehow disturbed by the bone conduction. The difference in the shape of the sounds when they were farther was not so well appreciated due to the low volume of the BC headset. Moreover, when we had several sounds, the lower ones were not perceived either clearly, so the participants had the impression of perceiving only the closest ones. Moreover, the perception was inversely correlated with the level (-.737, p = .037), but as almost all the participants chose the highest one, they faced the distance in the most difficult way. This correlation may be explained with the fact that in lower levels (such as the 3, with two vertical levels), the more quiet virtual ambient allowed the users to perceive subtle differences of distance (as in the pendulum example), undetected in higher levels. These higher levels acted as noisy sonifications.

In the case of the height, differences in pitches are easy to perceive (maybe not so easy to be understood naturally). Finally, the ease of horizontal and vertical localization was correlated with the age: -.626 (p = .097) and -.702 (p = .052) respectively. The first correlation can be due to the evolution of the hearing capabilities along the life. The second one, even stronger, may be correlated with another effect of aging: the difficulties to learn new codes and, thus, to take advance of the brain plasticity. The age also contributed to lower comprehension of the sounds in the different steps (with inverse correlations marginally significant). The only exception to this rule is found in the last test, where the improvement in the comprehension is not correlated with the age of the participants. Moreover, in this test, the average increase of understanding is 4. This can be explained with the following reason: the system did not work properly inside the different houses. Additionally, the second test was very hard, with small objects on the table. Outside, the system used to work better, and the sounds were more meaningful.

Another interesting correlation found (although cannot be taken literally, see the limitations, at the end of this section) shows a correspondence between the educational level and the confusion generated (Pearson’s correlation of -.657, p = .077), so that the higher the educational level, the lower the confusion is. This is coherent to the results obtained at the GeorgiaTech and shown in section 7.1.3.2.

Likewise, it is interesting to remark that the average comfort level chosen was 4.37, so most of participants chose the level 5, even though it is the most complex and difficult one.

The ease of the pose estimation was not perceived as high. This can be due to the difficultness to differentiate small changes in the height of the person, and more evident differences (like when sitting and standing up) are not assessed in the same way. This hypothesis is coherent with the error measured in the software evaluation and presented in figure 7.8. Most of the errors are produced in subtle differences, and not with the wider ones.

These especial difficulties to find small objects are also present in the table 7.6. Walls, cars or even people seem easier to be detected, meanwhile scaffolds, fences or mailboxes aren’t. The only exception in the order of the list is the trees, whose detections rates should be much higher.

Concerning the hardware itself, the form and weight was found to be adequate (4.1), not slow (2.5), and not bothering the glasses and BC headsets (2.5). Specifically, the BC seems to be a good solution to compatible the real and synthetic sounds in a very high degree (4.9), and it is perceived as an acceptable solution (4.5).

On the other hand, at the current point of development, the system is useful neither in indoors (2.75), nor in outdoors (2.75). The feeling of security decreased after using the system (2.9) and, in general terms, it was not perceived as useful in mobility (2.63). To understand these apparently contradictory answers, we have to focus on the list of reasons given by the participants in the different open questions.

The most cited problem is related with the algorithm of sonification. This part of the system should generate more and more different sounds to help perceiving the distance. This problem is technically related with another reason given 7 times: the lack of enough loudness, and the need of a control about it. But also new shapes and parameters of the sounds may help in this task.

The second most important problem was the false positives (mostly in higher frequencies), which impede, sometimes, to listen to “real sonifications” (i.e. sounds generated by the presence of real objects). In other occasions, it is just bothering and confusing, and the users get the idea that the system is unreliable. One of the participants stated that these errors may disable the whole system as a mobility aid.

The next most cited comment is again related with the sonification: the reduced level of panning. Most of the users complained about the difficulties to understand whenever a sound (i.e. an object) is on the right or in the left of the scene. This fact has already been discussed previously.

As it can be appreciated in the quantitative data as well, the loudness achievable by the system is not high enough, presenting problems in noisy environments.

An interesting suggestion of some participants was to design different functions, like “finding a door”, “finding a kiosk”, “finding luminosity”, “recognizing colors (maybe also in a semaphore)” or an optical character recognition. This increases the complexity of the system, but not equally in every case. Recognizing colors is quite simple in the current state, since the cameras capture color images. The same happens with the recognition of luminosity. Other tasks, such as recognizing objects or characters are much more complex computationally, and we should check whether the Raspberry Pi is able to perform such tasks with real time constrictions.

Finally, the main limitations of this test can be summarized in the following points:

- The small size of the sample. Eight people, even though all of them blind, is useful to extract qualitative information but no strong quantitative data. It is hard to obtain statistically significant data in this kind of experiments, since they involve a lot of time in both experimenters and participants, so large samples are very rare. Thus, this study should be read as a qualitative discussion about the proposed device rather than as a statistically based validation of it.

- The short training time. The participants lasted 5 hours in average to finish the whole process. The training of such a complex representation of reality in a cross-modal way should be much longer, giving better results within a period of months.
- Another problem has been the different scenarios in which the system has been tested. Sometimes the wall of the participants' houses was textured, some others it was completely flat and untextured, so the system did not work in the same way always. Moreover, outside, sometimes it was sunny, and other days it was cloudy, even raining. These changes in the ambient light and shades caused differences in the functioning of the system.

As a final summary, we gather all the constraints given in chapter 4 to be compared with the results obtained with semaphore based colors.

Area	Design goal	Result SW	Results HW
Image processing	Memory <307.2KB	230.4KB	230.4KB*
	Processing rate > 24fps	32fps	10.1fps
	Accuracy > 75%	75.2%	75.2%
Sonification	Real-time design	Yes**	Yes**
	Processing rate > 24fps	44191fps	28.8fps
	Level based design	6 levels	5 levels
	Accurate descriptors	Yes (table 7.1)	Yes (table 7.1)
	Enough channel bandwidth	Yes (headphones)	Yes (BC)
	Boolean alarm	Yes	Yes
Global system	Baseline device	-	Yes (fig. 7.17)
	Autonomy > 3h	-	10h
	Intuitive training	VR/Experts***	Yes
	Cost < 500€	-	236€
	Global merit > 108e-3	-	322e-3

*Supposed images of 320×240. This system uses 170×120 images and, thus, the memory really used is 4 times less.

**With the exception of the vibrato which, however, works in real time mode.

***Intuitive training only with experts, but not intuitive –as described- when training the participants (VR).

Table 7.8. Summary of design goals and results obtained.

In this table we can appreciate one violation, and some participants complained about the low speed of processing of the HW system.

Finally, we can assume that hypothesis H1 and H3 have been proved to be true, although with some limitations. These limitations are more constraining in the case of the HW version working in outside environments, where the sounds, and the system as a whole, have posed some problems to the participants. Anyway, we should make the difference between the real prototype and the potential power of these APs and strategies.

Regarding hypothesis H4, the Raspberry Pi has shown to be a powerful and cheap device, which allows tiny systems (participants saw the real device as light and portable), however its computational power seems to be somehow limited and, for example, the images have been reduced by 4 to let the RPi process them in almost real-time. Moreover, more accurate algorithms need more computational power, so given that this was one of the most important problems found, we can affirm that technology still needs some time to achieve power and cheap enough commercial implementations.

We can summarize the hypothesis and proofs in the following table.

Research Hypothesis	Comprobation
H1: Technology can help visually impaired people to move securely.	VR tests (section 6.6.1.2), real system (7.1.3.2 and 7.2.4.2) and in general terms, in 6.8 and 7.2.5)
H2: Light and fast image processing algorithms can be designed, and can run over cheap and generic hardware.	1st statement in sections 5.4.3, 5.5.5 and 5.6. 2nd, with limitations, in 7.2.4.1
H3: Sounds can substitute mobility relevant aspects of the visual world (the so-called *sonification*).	Sections 6.6.1.2, 6.8, 7.1.3.2, 7.2.4.2 and 7.2.5
H4: Commercial technology is mature enough to implement a functional and low-cost AP.	Raspberry Pi, with limitations discussed in 7.2.5

Table 7.9. Summary of hypothesis and comprobations.

Please watch this video recorded after the final tests.

References

1. ICECAT, "NGS NETCam300." *http://icecat.es/p/ngs/netcam-300/webcams-8436001305400-netcam300-3943712.html* . 2013.

2. P. Revuelta Sanz, B. Ruiz Mezcua, and J. M. Sánchez Pena, "Users and Experts Regarding Orientation and Mobility Assistive Technology for the Blinds: a Sight from the Other Side." *Proceedings of the AIRTech Int.Conference 2011.* 2011.

3. D. Bavelier and H. J. Neville, "Cross-modal plasticity: where and how?" *Nature Reviews Neuroscience* vol. 3 no. 443, p.452. 2002.

4. L. i Stefano, M. Marchionni, and S. MatToccia, "A fast area-based stereo matching algorithm." *Image and Vision Computing* vol. 22, pp. 983-1005. 2004.

5. Y. Jia, Y. Xu, W. Liu et al., "A Miniature Stereo Vision Machine for Real-Time Dense Depth Mapping." *Lecture Notes in Computer Science, Computer Vision Systems* vol. 2626, pp. 268-277. 2003. Berlin Heidelberg, Springer-Verlag.

6. G. Kraft and R. Kleihorst, "Computing Stereo-Vision in Video Real-Time with Low-Cost SIMD-Hardware." *Lecture Notes in Computer Science, Advanced Concepts for Intelligent Vision Systems* vol. 3708, pp. 697-704. 2005. Berlin Heidelberg, Springer-Verlag.

7. S.-K. Han, M.-H. Jeong, S.-H. Woo et al., "Architecture and Implementation of Real-Time Stereo Vision with Bilateral Background Subtraction." *Lecture Notes in Computer Science, Advanced Intelligent Computing Theories and Applications.With Aspects of Theoretical and Methodological Issues* vol. 4681, pp. 906-912. 2007. Berlin Heidelberg, Springer-Verlag.

8. S.-H. Lee, J. Yi, and J.-S. Kim, "Real-Time Stereo Vision on a Reconfigurable System." *Lecture Notes in Computer Science, Embedded Computer Systems: Architectures, Modeling, and Simulation* vol. 3553, pp. 299-307. 2005. Berlin Heidelberg, Springer-Verlag.

9. CYPRESS, "PSoC® 3 and PSoC 5LP USB Transfer Types." *http://www.cypress.com/?rID=39553* . 2013.

10. Silicon Lab.Inc., "Toolstick F330DC." *http://www.silabs.com/Support%20Documents/TechnicalDocs/ToolStick-F330-DC-UG.pdf* . 2013.

11. Sound Marketing Organization, "Audio Digital Signal Processors for Installed A/V Systems." . 2006.

12. Texas Instruments, "TMS320VC5506 Fixed-Point Digital Signal Processor." *http://docs-europe.electrocomponents.com/webdocs/0b47/0900766b80b47b69.pdf* . 2013.

13. Smart-GSM, "Características Técnicas HTC EVO 3D." *http://www.smart-gsm.com/moviles/htc-evo-3d* . 2012.

14. Geek, "$74 MK802 Android micro-PC beats Cotton Candy to the punch." *http://www.geek.com/articles/chips/74-mk802-android-micro-pc-beats-cotton-candy-to-the-punch-20120517/* . 2013.

15. Geek, "$89 dual-core Android stick PC leaves MK802 in the dust." *http://www.geek.com/articles/chips/89-dual-core-android-stick-pc-leaves-mk802-in-the-dust-20120814/* . 2013.

16. Geek, "Z902 Android micro-PC one-ups the MK802 for $75." *http://www.geek.com/articles/chips/z902-android-micro-pc-one-ups-the-mk802-for-75-2012087/* . 2013.

17. Aliexpress, "A02 Android Micro Computer Mini PC HDMI Stick 1080P FUll HD Video Android 4.0 ICS 512MB RAM 4GB." *http://www.aliexpress.com/store/product/Android-Micro-Computer-Mini-PC-HDMI-Stick-1080P-FUll-HD-video-Android-4-0-ICS-512MB/803232_580613871.html* . 2013.

18. RaspberryPi, "RaspberryPi Project." *http://www.raspberrypi.org/* . 2013.

19. Texas Instruments, "MSP430™ Ultra-Low Power 16-Bit Microcontrollers." *http://www.ti.com/lsds/ti/microcontroller/16-bit_msp430/overview.page?DCMP=MCU_other&HQS=msp430* . 2013.

20. Xilinx, "Spartan-3AN FPGA Family Data Sheet." *http://docs-europe.electrocomponents.com/webdocs/0db5/0900766b80db54aa.pdf* . 2009.

21. eLinux, "RPi Hardware." *http://elinux.org/RPi_Hardware* . 2013.

22. Computer Lab.Univ.of Cambridge, "Image Processing Library." *http://www.cl.cam.ac.uk/downloads/freshers/image_processing.tar.gz* . 2013.

23. Pd-community, "Pure Data." *http://puredata.info/* . 2013.

24. GameChanger Products LLC, "How It Works." *http://www.audioboneheadphones.com/howitworks.html* . 2008.

25. Planetronic, "Amplificador/Splitter de Audio portátil c/Bateria – 2xJack 3.5mm." *http://www.planetronic.es/amplificadorsplitter-audio-portatil-cbateria-2xjack-35mm-p-106724.html* . 2013.

26. Amazon, "Anker® Astro2 Dual USB Output 8400mAh Backup External Battery Pack Charger with built-in Flashlight for iPhone 5, all iPhone, iPad, iPod models; Android Smartphones: Samsung Galaxy S3 S III I9300, Galaxy S2 S II I9100, Galaxy Nexus, Galaxy Note / HTC Sensation, One X, One S, Thunderbolt, EVO / Nokia N9 lumia 900 800 / Motorola Triumph, Droid 3 X X2 Bionic Razr; PS Vita; GoPro; and More Mobile Devices [2 USB Output 5V 1A / 2A, Faster Charging." *http://www.amazon.com/Upgraded-Version-External-Flashlight-Smartphones/product-reviews/B0067UPRQ4* . 2013.

27. J. A. MacDonald, P. P. Henry, and T. R. Letowski, "Spatial Audio through a Bone Conduction Interface," *International Journal of Audiology,* vol. 45. pp.595-599, 2006.

28. B. N. Walker and R. M. Stanley, "Evaluation of Bone Conduction Headsets for use in Multitalker Communication Environments," *Proceedings of the Human and Ergonomics Society, 49th Annual Meeting.* pp.1615-1619, 2005.

8. Conclusions, Further Works and Contributions

In this work, different original contributions have been made to the state-of-the-art of Assistive Technologies. Firstly, several local and detailed evaluations have been presented in each chapter in order to evaluate the performance of the different subsystems designed and implemented. In the chapter 7, the whole system was presented and evaluated by users under real operation condictions.

The most relevant conclusions will be presented, with a presentation and discussion of the further works needed to fix the main problems detected and improves the functionality of the devise.

Finally, a summary of the scientific publications obtained in the framework of this thesis is also presented.

8.1. Conclusions

The main objective of the research was to build a first prototype of a new assistive product, which has shown to be complex and assorted. In this path, contributions to different fields, as the image processing, the sonification, the assistive products and even in the field of political philosophy were carried out along this work.

Four main hypotheses were formulated in section 1.5:

- H1, stating that technology may help visually impaired people to move more secure, has been validated through the tests with the VR (presented in section 6.6.1.2), with the real system (discussed in sections 7.1.3.2 and 7.2.4.2) and, in general terms, in sections 6.8 and 7.2.5.
- H2 says that light and fast image processing algorithms can be designed, and can run over cheap and generic hardware, as has been proven in sections 5.4.3, 5.5.5 and 5.6 for the first assumption, and in section 7.2.4.1 for the second assumption, about the cheap and generic HW.
- H3 affirms that sounds can substitute mobility relevant aspects of the visual world. This has been found to be true, as explained in sections 6.6.1.2, 6.8, 7.1.3.2, 7.2.4.2 and 7.2.5.
- H4 states that technology is mature enough to implement functional and low-cost APs.

Results obtained with the Raspberry Pi partially confirm the hypothesis H4. In this case, the Raspberry Pi has shown to be a powerful and cheap HW sub-system, which allows tiny systems (participants saw the real device as light and portable), however its computational power seems to be somehow limited and, for example, the images have been reduced by 4 to let the RPi process them in almost real-time. Moreover, more accurate algorithms need more computational power, so given that this was one of the most important problems found (see next paragraphs), we can affirm that technology still needs some time to achieve power and cheap enough commercial implementations.

As a human-centered research, several and different interactions with experts, potential users and sighted and visually impaired participants were, as well, crucial and carried out. Interviews to set the user requirements, and trainings and tests of different parts of the system, to keep the design as useful as possible were done during almost each step of the design chain.

Attending specifically to the different fields involved in this research, in the image processing area several papers and one book chapter were published (see section 8.5), proposing novel ways to solve the depth estimation problem through stereo vision. Up to four different options were proposed, showing a progression in their performance, and the best option found in this research was the one used in the rest of the system. However, the errors present in the algorithm are still quite high, as perceived by the participants in all the real system tests.

In the sonification field, a new mapping with redundancies has been proposed and tested over several different environments: a complete set of tests over virtual reality were prepared and evaluated, as well as different approaches to the real system, from artificial and mobility points of view. These tests were done, partially, with the help and expertise of the Sonification Lab at the School of Psychology, GeorgiaTech. Some others were carried out in Madrid, with an inestimable collaboration of different blind volunteers, who evaluated the system in real life situations, walking at home and even in the street.

The main problems found in the different evaluations were related with the errors of the real system, and some problems in the perception of differences in the sonification. The first ones were perceived to be extremely relevant to the well functioning of the system, and fixing them seems to be mandatory to build a marketable product.

Important correlations were found between the sociocultural class and/or age with the ease of taking charge of the system.

Likewise, participants proposed several great ideas to improve the system, adding functionalities to make it more useful for potential users, as detailed in the further works section.

8.2. Further Works

The results obtained on the different tests and evaluations give us the main paths in which we should deepen to overcome the problems found. We can analyze and propose solutions independently in different fields or aspects of the system.

8.2.1. Image processing

The main problem encountered by the participants during the real life tests, as well as in the artificial vision tests with the SW version was produced by the errors of the depth estimation algorithm. More accurate results seem mandatory to allow using the device in really safeness. This can be done by implementing new algorithms in parallel, as well as post-processing filters in order to eliminate the streaky lines and other visual noises.

The distance rage control was seen as a problem as well. Users need to control when an object produces sounds and when it doesn't, depending on the goal and surroundings of this moment. The response curve should be variable as shown in the following graphs.

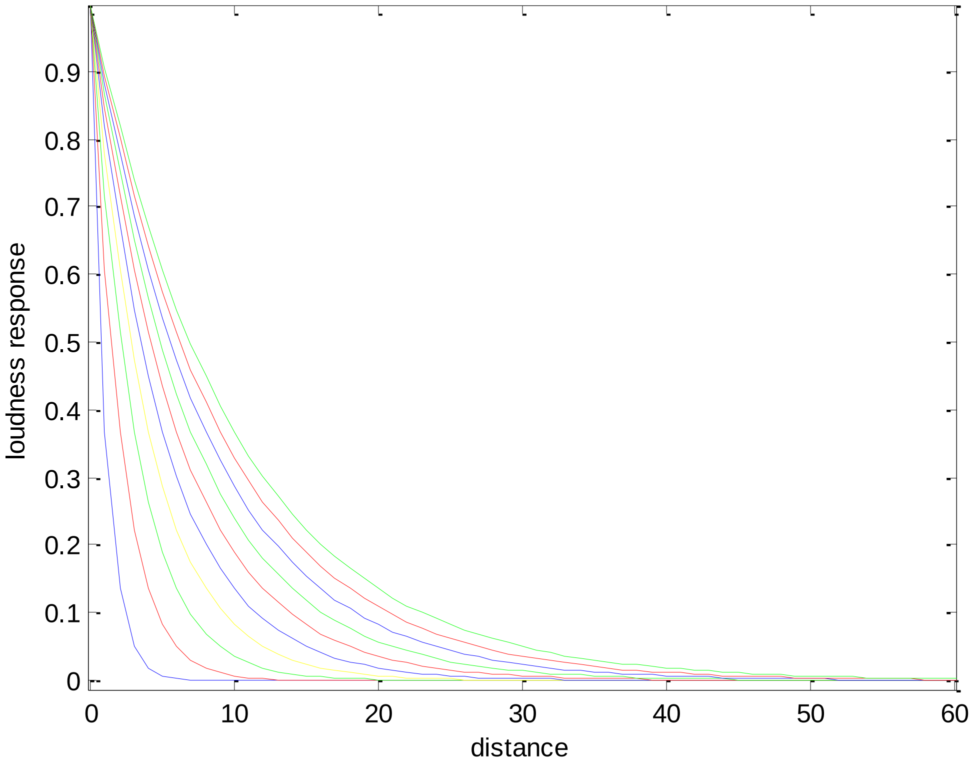


Figure 8.2. Different loudness response in terms of distance. These parametric curves should be selectable by the user.

Another interesting improvement, extracted from the participants' comments, is the adjustable area of interest. Again, this is not statically defined, because it depends on the task the user wants to do. Three paradigmatic approaches can be defined: searching mode, a "biscuit" and global mode. The first one, only should process a central area of the scene, so the attention and mental focusing on the object the user is searching for are not so high. The system would act like a pointer, with a narrow area of interest. In the second case, a biscuit-shape, vertical or horizontal, area of interest is sonified. In the third case, the whole image is sonified. The following figure shows these three paradigmatic approaches.

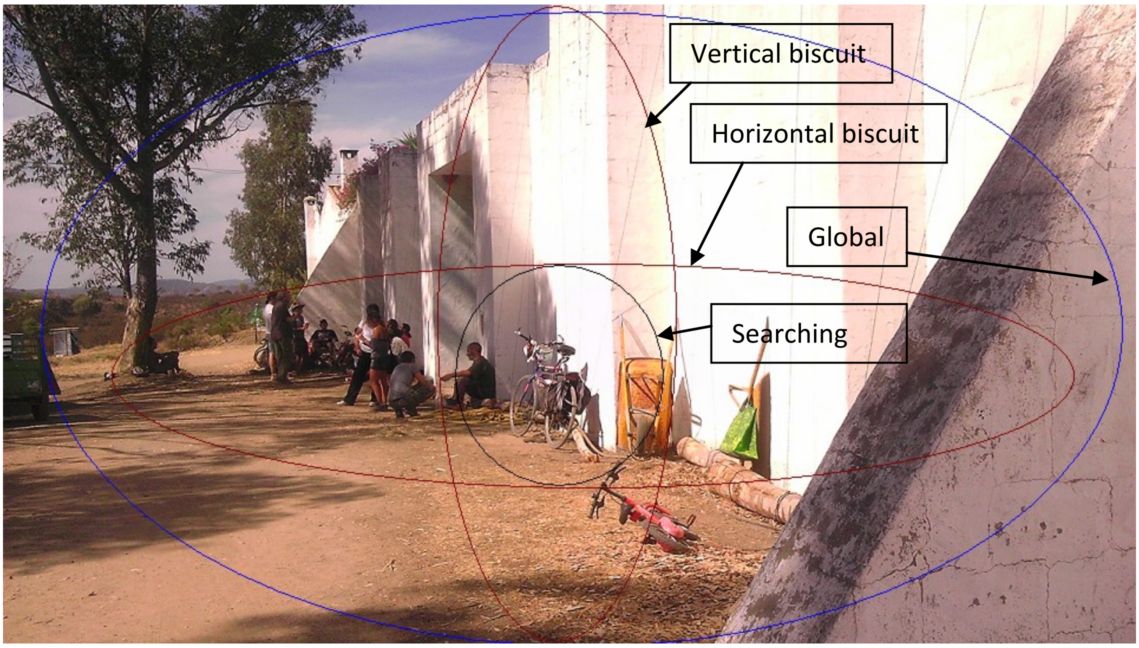


Figure 8.3. Three main areas of interest (two for the biscuit-shape) of a real life mobility scene.

Finally, several specific processing options have been proposed, to help in specific tasks:

- Color option: Blind people have evident problems when they want to know the color of a tissue, a paper or whatever other objects. There are tiny and useful systems to verbalize the color of an object. This option can easily be implemented in the ATAD, given that the cameras, nowadays, work in full RGB color map, and only a HSV transform should be needed to calculate the dominant color in some specific area of the image, for example, the center. This could be done as an extra function, pressing a button to hear a voice saying the color of the central area. This option would be helpful in mobility, as stated, when looking for shops, for example.
- In the same way, and due to the same reasons, the intensity of light is useful in some situations, as turning off artificial lights, finding the window, etc. This, again, only supposes a minimal change in the code to sonify, as done by other devices, the intensity of light in terms of pitch, loudness or other parameter. Again, only the central area of the image should be sonified, to avoid distortions in complex scenes.
- Another user proposed a differential function, i.e. something to know when the scene has changed, to find doors or pictures in a wall, differences of textures, etc. A combination of both spatial and temporal differential filters could be implemented as specific options. The utility and performance of each one of them should be evaluated to know which one is more useful, or which combination of them is the optimum implementation.
- The next step can be, as also proposed, the object recognition. Although this enters in more complex computer vision algorithms, it is obvious that this would simplify the mobility of the users. This has been already heard in the initial interviews, and again at the final evaluation. Even if simple objects are not that hard to be detected (as semaphores or trees, walls, etc.), more complex ones, as rubbish, scaffolds, tables or

persons need much more processing power, since they are more sensitive to perspective changes.

- Finally, but even more complex, the OCR option has also been proposed. This is mandatory in a long term mobility aid, but still need more computational power and we are still a bit far from this.

8.2.2. Sonification

The sonification presented different problems or constraints that could be solved, improving the global usability of the system. Most of them have been proposed by the participants and some others are provided by the author.

Distance representation is one of the most important aspects of the mobility. Thus, when it is not clear enough, it becomes the most important limitation of the system. Participants stated that more clear sounds should be needed to help in the discrimination of distances, since loudness and shape of the sound were not clear enough. The low pass filter could be redefined to have a more abrupt behaving against the distance, and the loudness could be varied in a more violent way as well. One of the participants also proposed an interesting solution: to implement a frequency beep as done in the cars' parking systems. The sonification, following this rule, would produce the same sounds, but when the objects are far, their sounds would only appear from time to time, and when the object comes closer, the intermittency of the sounds becomes smaller till the alarm level, when the sounds are stable, as they are now independently of the distance. This is not exactly real time sonification, but it could work very well for mobility issues, with the addition advantage that far objects bother even less since their sounds are only sonified with long intermittencies.

The panning was also perceived as deficient. Participants stated that they had problems to understand where the object was in the horizontal axis. More forced panning can be implemented, although not virtually locating it where it is not (i.e., if the object only produces sounds in one ear, this is not only out of the visual field, but it is completely unreal, since no real sound source can only excite one ear).

Related with the previous complain, some participants proposed as well to invert the vibrato. This is a meaningful idea. The vibrato seemed to be more easily perceived, so they found a paradox: the objects in the center, the most dangerous ones, are perceived more hardly than the lateral ones, less relevant to the mobility. Thus, an inversion of the vibration, to make central objects vibrate and lateral objects flatter could help in the understanding of the danger.

Another problem has been invisible for every participant. Following the design of the sonification, explained in both chapters 6 and 7, the maximum loudness of the image sonified depends of the absolute intensity of the pixels sonified. This was very simple to implement, but generate important problems:

- Higher levels (with higher number of pixels sonified) produce higher loudness.

- An object just occupying one pixel (for example, it is entering in the scene from a lateral or vertically) produce a lower sound that and object occupying several of them. Paradoxically, a farther and bigger object could produce higher loudness than a smaller and closer one, much more dangerous.

These two main problems derived from the design should be fixed. For that, the system should weight the image, and apply a standard loudness depending on the distance, independent of the number of pixels of such object (or distance, no need to understand that different pixels belong or not to the same object). Likewise, every level should have the same maximum loudness, so this information can be extrapolated from one level to the following one.

Finally, some participants asked for a verbal level, where the system only produces verbal messages about the mobility, such as “free way”, “obstacle at your right”, etc. This could be taken into account and easily implemented in a complete system as the RPi, where real-voice recorded messages have, indeed, already used to know the internal state of the system during the tests.

8.2.3. Hardware

Many different aspects of the HW version have been partially criticized and, hence, some solutions should be proposed.

The first one focuses on the lack of computational power to achieve better depth estimations. In this case, two RPi can be attached, deriving different processing tasks to each one so we can multiply the total computational power and, thus, implement slower but more accurate algorithms. The discussion should be then focused on which tasks can be running in parallel and how both devices may communicate without converting this last point in the bottleneck of the system.

Regarding with this option, we have found double core microcomputers running in Android. We already saw the problems of synthesizing real time MIDI messages under Android, but we could use these microcomputers just for the image processing, letting the sound generation to specifically designed hardware. This options should give much better performances, although the price would increase a significantly (maybe even the double).

An important problem found with the real images, in both the SW and the HW versions, were the images capture. The cameras used, two commercial webcams, have their own control of the exposure and white balance. This produced important problems in the depth estimation, when the image in one of them was much brighter than in the other one, so the exposure changed dramatically in one of them, and the algorithm wasn’t able anymore to match the pixels in the other one. This must be SW controlled, or at least one of them working in this automatic mode and the other one in a slave mode, to keep the same exposure and white balance in both captured images. The cameras, finally, could be smaller (as those of the cell phones) to be embedded behind standard glasses.

The autonomy has been found to be so high, overall because extra hours are correlated with extra size and weight. We could propose smaller batteries, sacrificing some hours of autonomy, to offer a smaller (and cheaper) product. Moreover, the used battery in the prototype was a commercial cell phone charger, with its own Li-Ion to USB converter. This can be easily designed and, thus, only needing a single battery, much cheaper and smaller.

The loudness is a problem itself, despite the specific sonification problem related to this and already discussed. Both a loudness control and an Automatic Gain Control (AGC) should be implemented in the device, to allow higher flexibilities in different ambient. The loudness control could control the maximum (or average, it's irrelevant) AGC loudness and smaller variations automatically produced by the system in correlation with the ambient sound. A noise sensor should be included in the system to sense the ambient sound level.

The user needs to control the device immediately and without rebooting, as it is the case now. Moreover, the level choice is currently done with an external numeric keyboard, which must be released after the selection and then, the second camera is attached. A button to control the profile level is mandatory so the user can change as many times as s/he wants, the profile level depending on the situation or her/his own willing.

The system has also seen as transformable to a pointer. This helps in scanning the ambient without moving the head so much, something not habitual for the blinds. Thus, the cameras could be changed from the glasses to a pointer (ant then, for example, attached to the cane).

This last option, however, could be compatible with the next problem detected and the proposed solution. The cameras (attached to the glasses) and the BC device should be one. It was found to be bothering having two different devices in the head, and some BC glasses (suitable to attach the cameras to them) are available in the market. We should need to think how to make compatible both requirements.

Another problem of the first prototype is the lack of a box including all the hardware together and secure. This will be the last step, given that the HW can change due to the solutions proposed, although it must be integrated in a single piece.

Finally, one participant was thinking about how to help in the training, and proposed the design of an online videogame, so blind (and even sighted) people can try the system in virtual mobility ambient to get used to it, as well as to evaluate it before buying it.

8.3. Scientific Contributions of this Research

In this section, the list of publications related with the research performed in the frame of this Thesis work is given.

8.3.1. Book Chapters

- Authors: Revuelta Sanz, P., Ruiz Mezcua, B., & Sánchez Pena, J. M.
  Title: Depth Estimation. An Introduction.
  Book title: Current Advancements in Stereo Vision. 224 pp.
  Editor: Ms. Marina Kirincic. InTech Ed. In press.
  Place: Rijeka, Croatia.
  Date: 2012
  ISBN: 978-953-51-0660-9, ISBN 979-953-307-832-7
  http://www.intechopen.com/books/export/citation/EndNote/current-advancements-in-stereo-vision/depth-estimation-an-introduction

- Authors: Revuelta Sanz, P., Ruiz Mezcua, B., & Sánchez Pena, J. M.
  Title: ICTs for Orientation and Mobility for Blind People. A State of the Art.
  Book Title: ICTs for Healthcare and Social Services: Developments and Applications
  Editor: Isabel Maria Miranda & Maria Manuela Cruz-Cunha. IGI-Global.
  Place: Hershey, EE.UU
  Date: 2011.

8.3.2. Peer Reviewed International Conferences

- Authors: Pablo Revuelta Sanz, Belén Ruiz Mezcua, José M. Sánchez Pena
  Title: A Sonification Proposal for Safe Travels of Blind People.
  Conference: The 18th International Conference on Auditory Display (ICAD2012).
  Publication: (not available yet).
  Place: Atlanta, Georgia (U.S.A.).
  Date: June 18-22, 2012

- Authors: Pablo Revuelta Sanz, Belén Ruiz Mezcua, José M. Sánchez Pena
  Title: Sonification as Rights Implementation.
  Conference: The 18th International Conference on Auditory Display (ICAD2012).
  Publication: (not available yet).
  Place: Atlanta, Georgia (U.S.A.).
  Date: June 18-22, 2012

- Authors: Pablo Revuelta Sanz, Belén Ruiz Mezcua, José M. Sánchez Pena
  Title: Estimating Complexity of Algorithms as a Black-Box Problem: A Normalized Time Index.
  Conference: 3rd International Multi-Conference on Complexity, Informatics and Cybernetics (IMCIC 2012).
  Publication: (not available yet).
  Place: Orlando, Florida (U.S.A.).
  Date: March 25th-28th, 2012.

- Authors: Pablo Revuelta Sanz, Belén Ruiz Mezcua, José M. Sánchez Pena
  Title: Users and Experts Regarding Orientation and Mobility Assistive Technology for the Blinds: a Sight from the Other Side.
  Conference: AIRtech 2011: Accessibility, Inclusion and Rehabilitation using Information Technologies.
  Publication: Proceedings of the AIRTech, pp. 3-4.
  Place: La Habana (Cuba).
  Date: December 13-15, 2011
  Selected paper to be sent to the Journal of Research and Practice in Information Technology.

- Authors: Pablo Revuelta Sanz, Belén Ruiz Mezcua, José M. Sánchez Pena
  Title: A Review of Orientation Technologies for the Blinds
  Conference: AIRtech 2011: Accessibility, Inclusion and Rehabilitation using Information Technologies.
  Publication: Proceedings of the AIRTech, pp. 7-8.
  Place: La Habana (Cuba).
  Date: December 13-15, 2011
  Selected paper to be sent to the Journal of Research and Practice in Information Technology.

- Authors: Pablo Revuelta Sanz, Belén Ruiz Mezcua, José M. Sánchez Pena
  Title: A Review of Mobility Technologies for the Blinds
  Conference: AIRtech 2011: Accessibility, Inclusion and Rehabilitation using Information Technologies.
  Publication: Proceedings of the AIRTech, pp. 5-6.
  Place: La Habana (Cuba).
  Date: December 13-15, 2011
  Selected paper to be sent to the Journal of Research and Practice in Information Technology.

- Authors: Pablo Revuelta Sanz, Belén Ruiz Mezcua, José M. Sánchez Pena
  Title: Stereo Vision Matching over Single-Channel Color-Based Segmentation
  Conference: International Conference on Signal Processing and Multimedia Applications (SIGMAP 2011)
  Publication: Proceedings SIGMAP 2011, pp. 126-130.
  Place: Seville (Spain).
  Date: July, 2011.

- Authors: Pablo Revuelta Sanz, Belén Ruiz Mezcua, José M. Sánchez Pena
  Title: ATAD: una Ayuda Técnica para la Autonomía en el Desplazamiento. Presentación del Proyecto
  Conference: IV Congreso Internacional de Diseño, Redes de Investigación y Tecnología para todos (DRT4ALL 2011)

Publication: Libro de Actas DRT4ALL 2011, pp. 151-161.
Place: Madrid (Spain).
Date: June, 2011

- Authors: Pablo Revuelta Sanz, Belén Ruiz Mezcua, José M. Sánchez Pena
  Title: Efficient Characteristics Vector Extraction Algorithm using Auto-seeded Region-Growing.
  Conference: 9th IEEE/ACIS International Conference on Computer and Information Science (ICIS 2010)
  Publication: Proceedings of the 9th IEEE/ACIS International Conference on Computer and Information Science (ICIS 2010), pp. 215-221.
  Place: Kaminoyama (Japan).
  Date: August, 2010.

8.3.3. Peer Reviewed Journals (Indexed JCR)

- Authors: Pablo Revuelta, Belén Ruiz Mezcua, José Manuel Sánchez Pena, Bruce N. Walker.
  Title: Scenes and images into sounds: a taxonomy of image sonification methods for mobility applications
  Journal: Journal of the Audio Engineering Society (in press)
  Date: March 2013.

- Authors: Pablo Revuelta, Belén Ruiz Mezcua, José Manuel Sánchez Pena, Jean-Phillippe Thiran.
  Title: Segment-Based Real-Time Stereo Vision Matching using Characteristics Vectors
  Journal: Journal of Imaging Science and Technology 55(5)
  Date: Sept./Oct. 2011.
  Note. This paper was selected as “feature article” of the issue and top 20 most downloaded in December 2011.

8.3.4. Papers under Review

- Authors: Pablo Revuelta, Belén Ruiz Mezcua, José Manuel Sánchez Pena
  Title: Fast and Dense Depth Map Estimation for Stereovision Low-cost Systems
  Journal: Image Science Journal
  Date: Feb. 2013.

- Authors: Pablo Revuelta, Belén Ruiz Mezcua, José Manuel Sánchez Pena, Bruce N. Walker.
  Title: Evaluation of a Mobility Assistive Product for the Visually Impaired
  Journal: IEEE Transactions on Rehabilitation
  Date: June. 2013

- Authors: Pablo Revuelta, Belén Ruiz Mezcua, José Manuel Sánchez Pena, Bruce N. Walker.
  Title: Evaluation of an Artificial Vision System for the Visually Impaired
  Journal: Sensors
  Date: May. 2013
  Provisionally Accepted.

- Authors: Pablo Revuelta, Belén Ruiz Mezcua, José Manuel Sánchez Pena.
  Title: Implementation of a Technical Aid for an Autonomous Displacement (ATAD)
  Conference: ICAVITECH 2013.
  Date: May. 2013.

8.4. Other Scientific Merits

8.4.1. International stays

- May 2012 – July 2012: **pre-Ph.D. scholar in the Georgia Institute of Technology (GeorgiaTech)**, Sonification Lab, School of Psychology, Atlanta, U.S.A.

- June 2010 – August 2010: **pre-Ph.D. scholar in the École Polytechnique Fédéral de Lausanne (EPFL)**, Digital Signal Processing Laboratory 5, Lausanne, Switzerland.

8.4.2. Other Publications

- Authors: J. Jiménez, P. Revuelta, L. Moreno and A. Iglesias
  Title: Evaluating the Use of Speech Technologies in the Classroom: The APEINTA Project.
  Conference: World Conf on Educational Multimedia, Hypermedia and Telecommunications
  Publication: Proc.s of World Conf on Educational Multimedia, Hypermedia and Telecommunications pp. 3976-3980.
  Place: Toronto (Canada).
  Date: July, 2010.

- Authors: Iglesias, L. Moreno, J. Jiménez and P. Revuelta.
  Title: Evaluating the Users' Satisfaction Using Inclusive Initiatives in Two Different Environments: The University and a Research Conference.
  Conference: 12th International Conference on Computers Helping People with Special Needs.
  Publication: Proc.s of Computers Helping People with Special Needs. vol. 6179/2010, pp. 591-594. 2010.
  Place: Viena (Austria).
  Date: May, 2010.

- Authors: Pablo Revuelta Sanz, Javier Jiménez Dorado, Ana Iglesias, Lourdes Moreno
  Title: APEINTA: a Spanish educational project aiming for inclusive education in and out the classroom
  Conference: 14th ACM–SIGCSE Annual Conference on Innovation and Technology in Computer Science Education.
  Publication: ITICSE 2009: Proceeding of the 2009 Acm Sigse Annual Conference on Innovation and Technology in Computer Science Education, p. 393.
  Place: Paris (France).
  Date: July, 2009.

- Authors: Pablo Revuelta Sanz, Javier Jiménez Dorado, Ana Iglesias, Lourdes Moreno
  Title: APEINTA: un proyecto educativo español que apuesta por la inclusión dentro y fuera de las aulas
  Conference: AMADIS'09.
  Publication: Not available yet.
  Place: Barcelona (Spain).
  Date: June, 2009.

- Authors: Pablo Revuelta Sanz, Javier Jiménez Dorado, J. M. Sánchez Pena, Belén Ruiz Mezcua
  Title: Multiplatform System with Accessible Interfaces for Hearing Impaired Students.
  Conference: International Conference of Education, Research and Innovation 2008.
  Publication: Proceedings of the International Conference of Education, Research and Innovation 2008, p. 153.
  Place: Madrid (Spain).
  Date: November, 2008.

- Authors: Pablo Revuelta Sanz, Javier Jiménez Dorado, J. M. Sánchez Pena, Belén Ruiz Mezcua
  Title: Multidevice System fir Educational Accesibility of Hearing-impaired Students
  Conference: 11th IASTED International Conference on Computers and Advanced Technology in Education.
  Publication: ISSN: 978-0-88986-767-3, pp. 20-25. ACTA Press, Zurich.
  Place: Crete (Greece).
  Date: October, 2008.

- Authors: Pablo Revuelta Sanz, Javier Jiménez Dorado, J. M. Sánchez Pena, Belén Ruiz Mezcua
  Title: Subtitulado Cerrado para la Accesibilidad de Personas con Discapacidad Auditiva en Entornos Educativos
  Conference: SEPLN'2008
  Publication: Procesamiento del Lenguaje Natural 2008, pp. 305-306.
  Place: Madrid (Spain).

Date: September, 2008.

- Authors: Pablo Revuelta Sanz, Javier Jiménez Dorado, J. M. Sánchez Pena, Belén Ruiz Mezcua
  Title: Online Captioning System for Educational Resources Accessibility of Hard-of-Hearing people
  Conference: 11th International Conference on Computers Helping People with Special Needs, Proceedings Young Researchers.
  Publication: ICCHP Proceedings pp. 22-35.
  Place: Linz (Austria).
  Date: July, 2008.

- Authors: Jiménez Dorado, J., Revuelta Sanz, P., Ruiz Mezcua, B., & Sánchez Pena, J. M.
  Title: Recursos educativos accesibles en tiempo real para personas con discapacidad auditiva severa.
  Conference: AMADIS'08
  Publication: ISBN: 978-84-692-2811-4 AMADIS08, III Congreso de Accesibilidad a los Medios de Audiovisuales para Personas con Discapacidad, pp. 31-41.
  Place: Barcelona (Spain).
  Date: June, 2008.

- Authors: Revuelta Sanz, P., Jiménez Dorado, J., Ruiz Mezcua, B. & Sánchez Pena, J. M.
  Title: Automatic Speech Recognition to Enhance Learning for Disabled Students, (Chapter 7, pp. 89-104)
  Book Title: Technology Enhanced Learning for People with Disabilities: Approaches and Applications. 350 pp.
  Editor: P. Ordoñez, Jing-Yuan Zhao, and Robert Tenysson. Igi Global.
  Place: Hershey, EE.UU.
  Date: 2010
  ISBN-10: 1615209239, ISBN-13: 978-1615209231
  http://www.igi-global.com/bookstore/chapter.aspx?TitleId=45504

- Authors: Ana Iglesias, Lourdes Moreno, Elena Castro, Paloma Martínez, Javier Jiménez, Pablo Revuelta, José Manuel Sánchez y Belén Ruiz
  Title: APEINTA: Apuesta Por la Enseñanza Inclusiva: uso de las Nuevas Tecnologías dentro y fuera del Aula .
  Journal: Revista FIAPAS.
  Date: 2010.

8.4.3. Patents

- APUNTA: 5146 M-005804/2012, date 27/07/2012. Owner: UC3M.

Authors: Juan Francisco López Panea, Diego Carrero Figueroa, María Belén Ruiz Mezcua, José Manuel Sánchez Pena, Ana María Iglesias Maqueda, Javier Jiménez Dorado, Pablo Revuelta Sanz.

8.4.4. Awards

- The prototype "Subtitle Glasses for Hearing-impaired" developed by Pablo Revuelta, Javier Jiménez, Belén Ruiz and José Manuel Sánchez Pena was chosen by the TIME journal (U.S.A.) as one of "The Best Inventions of the Year", in "Entertainment" category.
  Aditional information:
  http://www.time.com/time/specials/2007/article/0,28804,1677329_1678427_1678437,00.html

- The Asociación de Retinosis Pigmentaria de Navarra granted the ARPN 2008 award to the CESyA (UC3M), because of its labour in cinema accessibility for disabled people.

- The project "Accessibility System for Hard of Hearing People in Educational Environments" (Pablo Revuelta, Javier Jiménez, Belén Ruiz and José Manuel Sánchez Pena) received a special mention (Finalist) in the ACCESS IT Awards (London, July 18th, 2008, sponsored by Microsoft, e-ISOTIS and AbilityNet).

- The project "Multidevice System for Educational Accessibility of Hearing-Impaired Students" (Pablo Revuelta, Javier Jiménez, Belén Ruiz and José Manuel Sánchez Pena) was proposed as one of the 5 finalists in the 7th International Competition of Ph.D. Students on Research in Technology-Based Education Area, IASTED'08 (Crete, Greece).

- FIAPAS 2008 Award to the APEINTA Project. The author was a participant in this project.

9. Resumen del Trabajo Realizado

El trabajo presentado describe el diseño, implementación y validación de una nueva ayuda técnica orientada a la asistencia en la movilidad de personas con discapacidad visual. Este trabajo de tesis está redactado íntegramente en inglés para obtener la "mención internacional". A continuación se presentan los resultados y contribuciones más relevantes en castellano.

9.1. Introducción y Justificación del Proyecto

Una ayuda técnica (AT) es "cualquier producto (incluyendo dispositivos, equipos, instrumentos, tecnologías y software) producidos especialmente o disponibles en el mercado, para prevenir, compensar, controlar, reducir o neutralizar deficiencias, limitaciones en la actividad y restricciones en participación "[1]. Esto significa una aplicación específica de la tecnología para mejorar la calidad de vida de las personas. En una definición amplia, casi toda la tecnología podría ser tomada como AT, como automóviles, computadoras o casas. En la práctica, el concepto de AT se aplica a la tecnología aplicada para resolver problemas en las personas con algún tipo de discapacidad temporal o permanente.

La OMS afirma que una AT es "cualquier producto (incluyendo dispositivos, equipos, instrumentos y software), especialmente fabricados o disponibles en general, utilizada por o para las personas con discapacidad:

- para la participación
- para proteger, apoyar, capacitar, medida o sustituto para las funciones corporales/estructuras y actividades
- para prevenir deficiencias, limitaciones en la actividad o restricciones en la participación " [2].

El trabajo se ha centrado en las personas con discapacidad visual, y más específicamente, en los problemas de movilidad y desafíos a los que este colectivo tiene que enfrentarse cada vez que sale de los entornos conocidos.

"Ceguera" es una palabra muy común para designar la situación en la que la gente no puede ver. Pero la definición médica debe ser más precisa que eso.

El primer problema que nos encontramos en la búsqueda de una definición más precisa es el umbral entre la "ceguera" y "baja visión". Un informe de la Organización Mundial de la Salud (OMS) propone la siguiente definición de "baja visión"[3]:

> Una persona con baja visión es aquella que tiene una alteración de la función visual aún después de tratamiento y/o corrección refractiva estándar, y tiene una agudeza visual de menos de 6/18 de percepción de luz, o un campo visual de menos de 10 grados desde el punto de fijación, pero que usa, o es potencialmente capaz de usar la visión para la planificación y/o ejecución de una tarea.

En el mismo informe, se afirma que "la definición actual no hace una distinción entre los que tienen ceguera " irreversible " (sin percepción de luz) y los que tienen percepción de luz pero siguen siendo inferiores a 3/60 en el mejor ojo "[3].

Por último, se puede encontrar una breve descripción de esta característica [4]:

> La ceguera es la incapacidad de ver. Las principales causas de ceguera crónica incluyen glaucoma catarata relacionada con la edad degeneración macular, opacidades corneales, la retinopatía diabética, el tracoma y las afecciones oculares infantiles (por ejemplo, causada por la deficiencia de vitamina A). Relacionada con la Edad ceguera está aumentando en todo el mundo, como es la ceguera debida a la diabetes no controlada. Por otro lado, la ceguera causada por la infección está disminuyendo, como resultado de la acción de salud pública. Tres cuartas partes de los casos de ceguera se pueden prevenir o tratar.

Del informe sobre la ceguera de la OMS [5], obtenemos la panorámica general siguiente:

- De los aproximadamente 314 millones de personas con discapacidad visual en todo el mundo, 45 millones son ciegos.
- La mayoría de las personas con discapacidad visual son mayores, y las mujeres están en mayor riesgo a cualquier edad, en cualquier parte del mundo.
- Cerca del 87% de la población con discapacidad visual vive en países en vías de desarrollo.
- El número de personas ciegas debido a enfermedades infecciosas ha disminuido mucho, pero la discapacidad relacionada con la edad va en aumento.
- Las cataratas siguen siendo la principal causa de ceguera en el mundo, excepto en los países más desarrollados.
- La corrección de errores de refracción podría dar una visión normal a más de 12 millones de niños (de edades comprendidas entre 5 y 15).
- Alrededor del 85% de todas las discapacidades visuales es evitable a nivel mundial.

Podemos encontrar en estos datos la magnitud de este problema, que afecta a casi el 3% de la población (en relación con las estimaciones de la OMS).

Un aspecto importante a destacar es que el deterioro visual genera dificultades para el movimiento o el acceso a la información, y esto puede generar una discapacidad si el medio ambiente no es compatible con las personas con estas dificultades. Así, la discapacidad visual es una combinación de algunos de los problemas fisiológicos y el entorno tecnológico, legal, económico y cultural.

Es muy importante señalar la diferencia entre la orientación y movilidad, a pesar de que suelen aparecer incluso en las mismas siglas como O&M. Siguiendo a los especialistas C. Martínez y K. Moss, *orientación* significa "saber dónde está en el espacio y dónde se quiere ir", mientras que la *movilidad* es "ser capaz de llevar a cabo un plan para llegar allí" [6]. Una clasificación similar fue propuesta por Petrie *et al.* [7] como "macronavegación" y "micronavegación", respectivamente. La primera definición apunta a una idea general de la situación en un sentido geográfico. La segunda tiene en cuenta cómo moverse o viajar, dada una posición y una dirección. Nos centraremos en estos trabajos sobre herramientas de movilidad, a las que se conoce generalmente como *Electronic Travel Aids* (ETAs).

Las personas ciegas pueden utilizar la información adicional proporcionada por estos dispositivos, junto con otra información ambiental, para complementar la escasa información visual, logrando una navegación exitosa [8].

Encontramos dos núcleos de motivación para proponer una nueva AT:

- Las personas ciegas (así como de otros colectivos con discapacidad) sufren la falta de derecho de la movilidad, en términos prácticos. Incluso si la arquitectura y el urbanismo ha implementado algunas soluciones, muchos obstáculos inesperados y cambios en el mobiliario urbano actúan como una barrera para un ejercicio efectivo de este derecho. Esto ha sido identificado hace tiempo como una de las limitaciones funcionales más importantes para las personas ciegas y con deficiencia visual [9, 10]. Este es un problema ético más que técnico, y sus fundamentos se discuten en detalle en [11], desde un punto de vista liberal.
- Se han encontrado problemas en el análisis de ATs, con respecto a facilidad de uso, la complejidad, el peso y el precio. Este campo debe seguir proponiendo soluciones a la movilidad de las personas ciegas.
- Se pretende proporcionar a las personas con discapacidad visual una AT que pueda ayudarles a incrementar su independencia en los desplazamientos de la vida diaria. Para ello, serán necesarios nuevos y ligeros algoritmos de procesamiento de imágenes, así como nuevas propuestas de sonificación, con el fin de reunir estos elementos hacerlos correr en hardware usable y barato.

Finalmente, tal y como se indica en [12], "los conceptos espaciales se utilizan para formar una comprensión conceptual, o un mapa cognitivo, de un entorno determinado".

La propuesta presentada en esta investigación tiene como objetivo "inducir" una imagen mental por medio de caminos no convencionales, ya que estos caminos están, de alguna manera, interrumpidos en las personas ciegas. Tenemos que trabajar con la llamada imaginería mental para ayudar a comprender el entorno.

Este proceso ha sido ampliamente estudiado. Los estudios iniciales que informan cómo funciona el cerebro cuando se trata de imágenes mentales comienzan en los primeros 80' [13-15], mientras que en términos psicológicos, los estudios se pueden encontrar desde mucho antes [16]. Estos primeros estudios proponen que las imágenes mentales se crean no sólo por los ojos, sino también a partir de otras fuentes. En [13], se ha informado que las imágenes mentales pueden representar relaciones espaciales que no estaban explícitamente en la formación de imágenes. Las podemos usar como cuando se escucha un plan de orientación o razonamiento espacial para formar tales imágenes mentales [13]. Además, las imágenes mentales responden a un punto de vista específico [13]. Por lo tanto, podemos suponer que un punto de vista subjetivo es intrínseco a la formación de imágenes mentales.

En los años siguientes, comenzamos a encontrar las primeras evidencias de las asimetrías hemisféricas y la especialización del cerebro en imágenes mentales [14, 15].

Los psicólogos han demostrado cómo funciona el cerebro para crear imágenes mentales con un poco más de material que el que viene de los ojos [17, 18].

Con estas técnicas, aparece una nueva hipótesis verificable empíricamente: "La imaginería visual es la capacidad de generar percepciones como las imágenes en la ausencia de entrada de la retina" [19]. Además, "las imágenes visuales pueden disociarse de la percepción visual" [[20]. Este es el efecto de la llamada " plasticidad cross-modal " del cerebro [21], la cual será explotada por el prototipo propuesto.

Las hipótesis globales son:

- H1: La tecnología puede ayudar a las personas con discapacidad visual a moverse con seguridad.
- H2: algoritmos de Luz y rápido procesamiento pueden ser diseñados, y se puede ejecutar sobre hardware barato y genérico.
- H3: Los sonidos pueden sustituir movilidad aspectos relevantes del mundo visual (la sonificación llamada).
- H4: La tecnología comercial es lo suficientemente madura para implementar un punto de acceso funcional y de bajo coste.

El objetivo último y global de este trabajo se centra en el diseño de un sistema de ayuda a la movilidad orientado a las personas con discapacidad visual, utilizando técnicas de procesamiento de imágenes y sonificación para detectar y representan obstáculos potencialmente peligrosos en ambientes interiores y exteriores. Para alcanzar el objetivo global, es necesario alcanzar los siguientes objetivos parciales:

- Objetivo 1: Procesamiento de imágenes: El procesador de imágenes será el sistema encargado de recuperar la información más relevante para la movilidad desde el mundo visual. Este es un elemento crucial del sistema, en cuanto a velocidad, precisión, la carga computacional y los sensores. Los objetivos principales de este subsistema serán un alto rendimiento, una baja tasa de error (para garantizar seguridad mínima), baja carga computacional y bajo costo del hardware necesario.
- Objetivo 2: Sonificación: El dispositivo utilizará sonidos para transmitir el entorno y la información relevante para la movilidad. El conjunto de sonidos debe ser claro, intuitivo y fácil de aprender. Se transmitirán dejando libre el oído para otros sonidos del mundo real, también importantes para un viaje seguro.
- Objetivo 3: Hardware: Teniendo en cuenta que el sistema debe ayudar en las tareas de movilidad, debe ser portátil, con un peso ligero y suficientemente autónomo para trabajar durante algunas horas seguidas. Por lo tanto, el hardware debe ser pequeño, y de baja potencia. Asimismo, y debido a otras razones, el hardware debe ser tan barato como sea posible, para ayudar en la difusión de la AT.

Por último, es necesario un estudio de usuarios en el que participen voluntarios con discapacidad visual y expertos para que el sistema propuesto muestre la eficacia de la solución implementada.

9.2. Revisión de las Ayudas Técnicas Disponibles

Para realizar el estado del arte en el campo de las ayudas técnicas a la movilidad y orientación de las personas ciegas, se ha revisado literatura técnica relacionada de la meta-base de datos Web of Science y las páginas web académicas o comerciales, cuando fue necesario.

Dicha búsqueda se ha dividido en dos fases, dedicadas a la movilidad y la orientación respectivamente. La primera búsqueda se llevó a cabo mediante las siguientes palabras clave: "ETA", "Electronic Travel Aid", "assistive product" y "assistive technology", junto con "blind" y "mobility".

La clasificación de las llamadas ETAs es compleja, ya que intervienen muchos parámetros, como la tecnología, la forma de uso, la información proporcionada, la aplicación espacial, etc.

En este estudio se clasifican las ETAs en cuanto al uso y la tecnología, y proporcionamos un ejemplo de cada subconjunto.

En el proceso de búsqueda se encontraron 80 productos o prototipos, de los cuales 17 de ellos son propuestos desde 1940 hasta 1970 y el resto entre 1970 y el presente (véase el tabla 9.1).

Uso	Nº ETAs	Tecnología	Ejemplo
Linterna	4	Infrarrojos	The UCSC Project [22]
		Ultrasonidos	The Polaron [23]
Bastón	13	Infrarrojos	Tom Pouce [24]
		Ultrasonidos	The Digital Ecosystem Sytem [25]
		Laser	The Laser Orientation Aid for Visually Impaired (LOAVI) [26]
Cinturón	7	Ultrasonidos	NavBelt [27]
Ropa	11	Visión Estéreo	Guelp Project “Haptic Glove” [28]
Cabeza	18	Visión Estéreo	Computer Aided System for Blind People (CASBliP) [29]
Externa	10	Infrarrojos	RIAS [30]

Tabla 9.1. Resumen de las ayudas técnicas para la movilidad.

Durante esta búsqueda, se encontraron muchos proyectos no comerciales que no están disponibles para la comunidad invidente, así como una explosión de este campo en los últimos 10 años. Este aumento de trabajos de investigación muestra lo importante que es, en la actualidad, la colaboración entre la tecnología y la asistencia social. Sin embargo, el mercado no permite la difusión adecuada de esta tecnología y, por lo tanto, la mayoría de ellos permanecen en un estado inutilizable.

El conjunto de los EOAs se pueden dividir en dos grupos principales, en relación con el entorno en el que se puede utilizar: interiores y exteriores. Asimismo, la forma de administración de la información al usuario permite otra clasificación, dependiendo del paradigma aplicado. En este trabajo se clasifican los EOAs en función del medio ambiente y la tecnología, y proporcionamos un ejemplo de cada subgrupo.

En el proceso de búsqueda se encontraron 34 productos de ayuda a la orientación, la mayoría de ellos diseñados para trabajar al aire libre con la tecnología GPS. El campo de investigación está limitado por el precio de la tecnología, y la democratización del GPS ha incrementado, desde el año 2000, las herramientas de orientación disponibles para las personas invidentes.

La forma de proporcionar información al usuario se limita a dar órdenes de voz sintética de dirección y el paradigma del "reloj", orientando al usuario por medio de una metáfora del reloj. Hay una excepción, el Body Mounted Vision System, que transmite por medio de ráfagas de tonos el error respecto al camino correcto.

La tabla 9.2 muestra la clasificación y algún ejemplo de cada subgrupo.

Uso	Nº EOAs	Tecnología	Ejemplo
Interiores	8	Balizas IR	The Cyber Crumbs [31]
		Balizas RFID	BIGS [32]
		Balizas Laser	The Instrumentation Cane [33]
		GSM	PYOM [34]
		Basado en PC	Subway Mobility Assistance Tool [35]
Exteriores	19	Mapas táctiles	NOMAD [36]
		GPS	BrailleNote GPS [37]
		Brújula	The University of Osnabrück Project [38]
Mixtos	7	Balizas IR	The Easy Walker [39]
		GPS+Bluetooth	Indoor Navigation System [40]
		Procesado de imagen	Body Mounted Vision System [41]

Tabla 9.2. Resumen de las ayudas técnicas para la orientación.

9.3. Requisitos de Usuarios

La norma de regulación ISO 13407 "Diseño de Procesos Centrados en Humanos para Sistemas Interactivos" [42], propone la inclusión de los usuarios potenciales desde los primeros pasos de desarrollo de cualquier proyecto. Sin embargo, esta recomendación no siempre se tiene en cuenta. Para incluir a las personas ciegas, se realizó una serie de entrevistas cualitativas con los usuarios potenciales de las AT con invidentes y/o expertas/os en campos relacionados, con el fin de describir cómo perciben estos colectivos la tecnología de asistencia, principales ayudas y dispositivos ya propuestos, críticas a los mismos y recomendaciones de diseño para lo que proponemos.

En el desarrollo de este estudio, se realizaron 11 entrevistas a diferentes perfiles (profesionales y personales) relacionados con la ceguera, la rehabilitación, la psicoacústica, la informática y la música.

Más en detalle, el espacio muestral de expertos entrevistados puede clasificarse en base a la siguiente categorización no exclusiva:

- Las personas ciegas: 6
- Psicología y rehabilitación perfil profesional: 2
- Perfil técnico profesional: 4
- Expertos en productos de apoyo: 5

- Expertos en música: 3

La entrevista presentó tres preguntas abiertas:

- Problemas por resolver en la vida cotidiana de la cortina, con respecto a la orientación y movilidad (pregunta formulada sólo para cegar los entrevistados y expertos en este campo).
- Los sistemas conocidos o dispositivos relacionados con estos problemas, y la crítica a los mismos.
- Propuestas y consejos acerca de nuevos sistemas (a nivel usuario o técnico).

Las respuestas se organizan siguiendo las preguntas. Los resultados obtenidos se detallan a continuación:

- Los principales problemas de la vida cotidiana están relacionados con conseguir orientarse en espacios desconocidos (estaciones de metro, sentirse "en el medio de la nada" y el problema de los ecos), con la falta de acceso a la información visual (falta de paneles en braille) y los obstáculos no detectables con el bastón o el perro-guía (bolardos, contenedores, buzones de correo, vallas de obra o andamios).
- Los usuarios no tienen un conocimiento profundo de las AT ya disponibles o propuestas, y sólo conocen algunas de ellas. Las principales críticas a las AT conocidas están encabezadas por el precio, que es el principal obstáculo para su democratización. Además, no hay economías de escala en este mercado y el público es limitado. Otro problema que se percibe es quién se ocupa del mantenimiento del sistema. El peso es otro problema importante de la mayoría de las AT comerciales, así como la dificultad de uso, es decir, cuán complejo es el dispositivo y su uso: "Los usuarios se volvían locos con el Ultracane", dijo una entrevistada. Este parámetro está relacionado con un tiempo de entrenamiento largo, como es el caso del EAV [43] o el vOICe [44].
- Los consejos dados para una nueva AT se relacionan con las críticas anteriores: Bajo precio, posibilidad de integración en el bastón, resistente al agua, fácil de manejar (especialmente para las personas mayores), portátil, tener en cuenta a las personas que no oyen suficientemente bien, diferentes perfiles para diferentes capacidades cognitivas, complejidad moderada y la mayor funcionalidad posible, disponibilidad para usuarios potenciales del grupo más amplio posible y aparato no ostentoso. Después de todo, los usuarios reclaman no generar una expectativa poco realista en la presentación de nuevos dispositivos.

9.4. Propuesta

De acuerdo con el objetivo principal de este trabajo, la arquitectura debe tratar de cumplir con los requisitos del usuario (capítulo 3 y sección anterior), así como con otros aspectos relacionados con la ergonomía, facilidad de uso, precio, entre otros.

En la figura 9.1 se muestran los esquemas de flujo funcionales y de información del sistema propuesto.

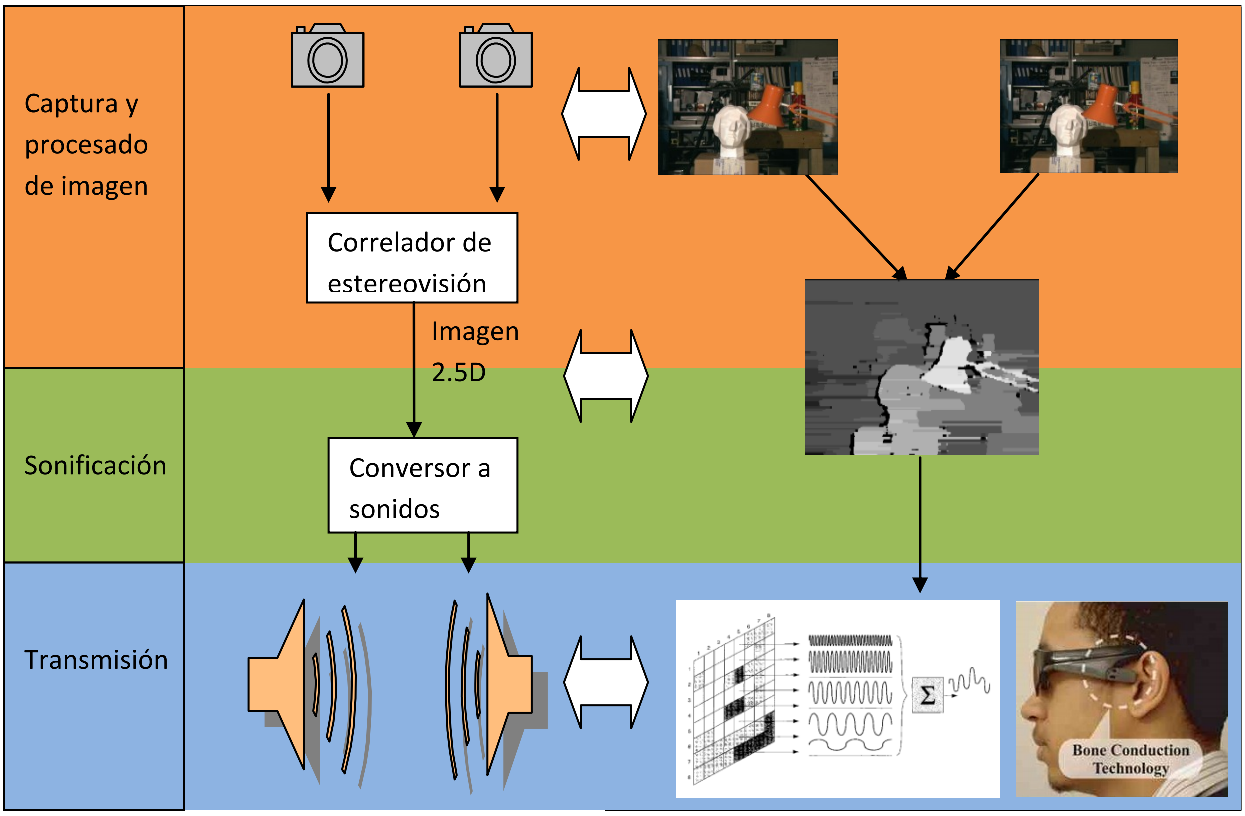

Captura y procesado de imagen
Sonificación
Transmisión
Correlador de estereovisión
Imagen 2.5D
Conversor a sonidos
Σ
Bone Conduction Technology

Fig. 9.1. Esquema funcional e informacional de la AT propuesta.

El sistema funciona de la siguiente manera:

- Dos microcámaras de bajo coste y comerciales capturan las imágenes, siguiendo el paradigma de la visión estereoscópica (véase el capítulo 5 o sección 9.5).
- Un correlador extrae el mapa de profundidad de la pareja de imágenes capturada y genera una imagen en formato 2.5D como salida. Esta imagen es una imagen en escala de grises con información sobre las distancias de cada pixel a las cámaras.
- Un sonificador, es decir, un bloque que convierte las imágenes a los sonidos, procesa la imagen 2.5D y genera un par adecuado de sonidos (dado que la sonificación propuesta es binaural).
- Finalmente, el transmisor envía la información acústica al usuario, por medio de un dispositivo de transmisión ósea.

Se han especificado algunos objetivos parciales para cada uno de los bloques y fases que componen el proyecto.

En cuanto al procesado de imagen, podemos resumirlos en tres puntos principales, proponiendo, además, algunas restricciones absolutas a estas variables:

- Baja complejidad computacional: El algoritmo de procesamiento de imagen debe ser lo más simple posible. El uso de procesador, dado que generalmente se relaciona con el coste, debe ser también lo más pequeño posible.
- Restricción de tiempo real (TR): Una condición importante del sistema de procesamiento de imágenes será procesar imágenes cercanas a condiciones de tiempo real, es decir, a 24 cuadros por segundo.

- Precisión > 75%: Con el fin de evitar errores de detección de objetos peligrosos, ni falsos positivos, se debe lograr una precisión de más del 75%.

En lo que concierne a la sonificación:

- Restricción de TR: Como se ha propuesto para el bloque de procesamiento de imágenes, la restricción de tiempo real es obligatoria a fin de proporcionar información actualizada y cambiante suavemente para el usuario. En efecto, este subsistema se debe sincronizar con el sistema de procesamiento de imagen para realizar una tarea única desde el punto de vista del usuario.
- Código intuitivo: Los sonidos utilizados para codificar la imagen debe aprovechar las capacidades auditivas naturales.
- Descriptores precisos: A pesar del objetivo anterior, se debe proporcionar información precisa al usuario, en cuanto a distancias y posiciones espaciales de los objetos y volúmenes detectados. La precisión no debe ser sacrificada absolutamente por la sencillez.

En cuanto a la transmisión de la información acústica, hay algunos objetivos importantes a tener en cuenta:

- Complejidad admitida. El canal a utilizar debe tener un ancho de banda lo suficientemente amplio. Esto significa que códigos por encima de alarmas booleanas deben poder ser transmitidos.
- Respeto al sistema auditivo. Dado que las personas ciegas valoran el sistema auditivo como su sentido más importante, el canal de transmisión propuesto debe respetar esta entrada la información, no suplantando los sonidos naturales del mundo.
- Alarma booleana. Independientemente del canal, una alarma de tipo booleana para peligros inminentes siempre debe poder ser aplicada.

En cuanto al sistema final, podemos proponer una autonomía mínima del dispositivo completamente funcional (procesando imágenes y sonificando) de alrededor de 3 horas. Pocos desplazamientos en la vida cotidiana supera este tiempo.

Finalmente, se puede definir una relación entre la velocidad, la precisión, la memoria utilizada y el precio que se debe maximizar, teniendo en cuenta algunas limitaciones absolutas. Esta relación se muestra en la siguiente expresión:

$$e = \frac{velocidad \cdot precisión \cdot autonomia}{precio} = autonomía \cdot (retardo \cdot errores \cdot precio)^{-1} \quad (9.1)$$

El parámetro de "mérito" global del sistema "e" será directamente proporcional a la velocidad y precisión, e inversamente proporcional al precio. La eq. 9.1 ignora la relevancia de memoria, ya que su contribución a la viabilidad del sistema no importante.

Podemos, por tanto, establecer un límite para el precio de 500 €, por lo que el mérito del sistema, en el caso pero, será:

$$e = \frac{24 \cdot 0.75 \cdot 3}{500} = 108 \cdot 10^{-3} \quad (9.2)$$

En cuanto al entrenamiento y usabilidad:

- Diferentes niveles/perfiles: Permitir que cada usuario pueda utilizar diferentes perfiles y niveles, desde los más simples a los interfaces más complejos (ver figuras 4.5 y 4.6).
- Explotación de la intuición: Combinado con el objetivo anterior, la formación debería aprovechar la intuición, para facilitar el entrenamiento y, por lo tanto, el uso de la AT.
- Basado en la vida cotidiana: Teniendo en cuenta que el entorno diario es familiar para todo el mundo, estos escenarios se deben utilizar como entornos de formación, lo que podría ayudar a las personas ciegas a incorporar el proceso de sonificación en los mecanismos inconscientes del cerebro.

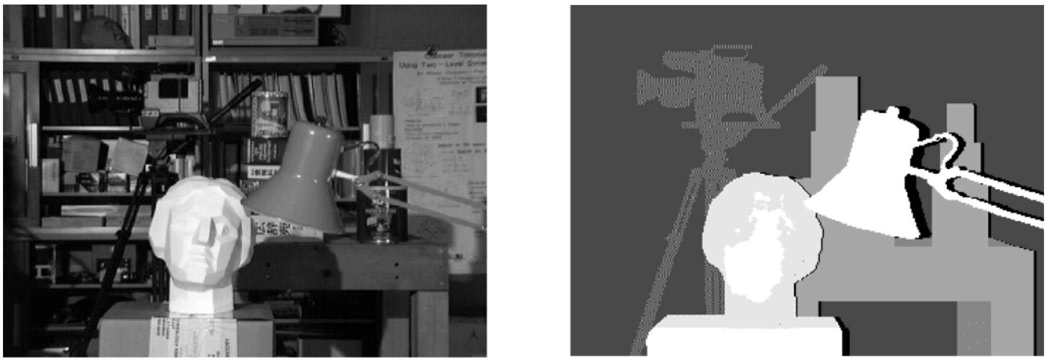
9.5. Procesado de Imagen

El capítulo 5, dedicado al procesado de imágenes, cubre un bloque fundamental de la AT propuesta.

Existen muchas y muy diversas propuestas, técnicas, paradigmas y algoritmos para resolver la estimación de la profundidad detalladas, por ejemplo, en [45]. Tras una revisión de dichos sistemas, se optó por utilizar el paradigma de visión estéreo por diversas razones:

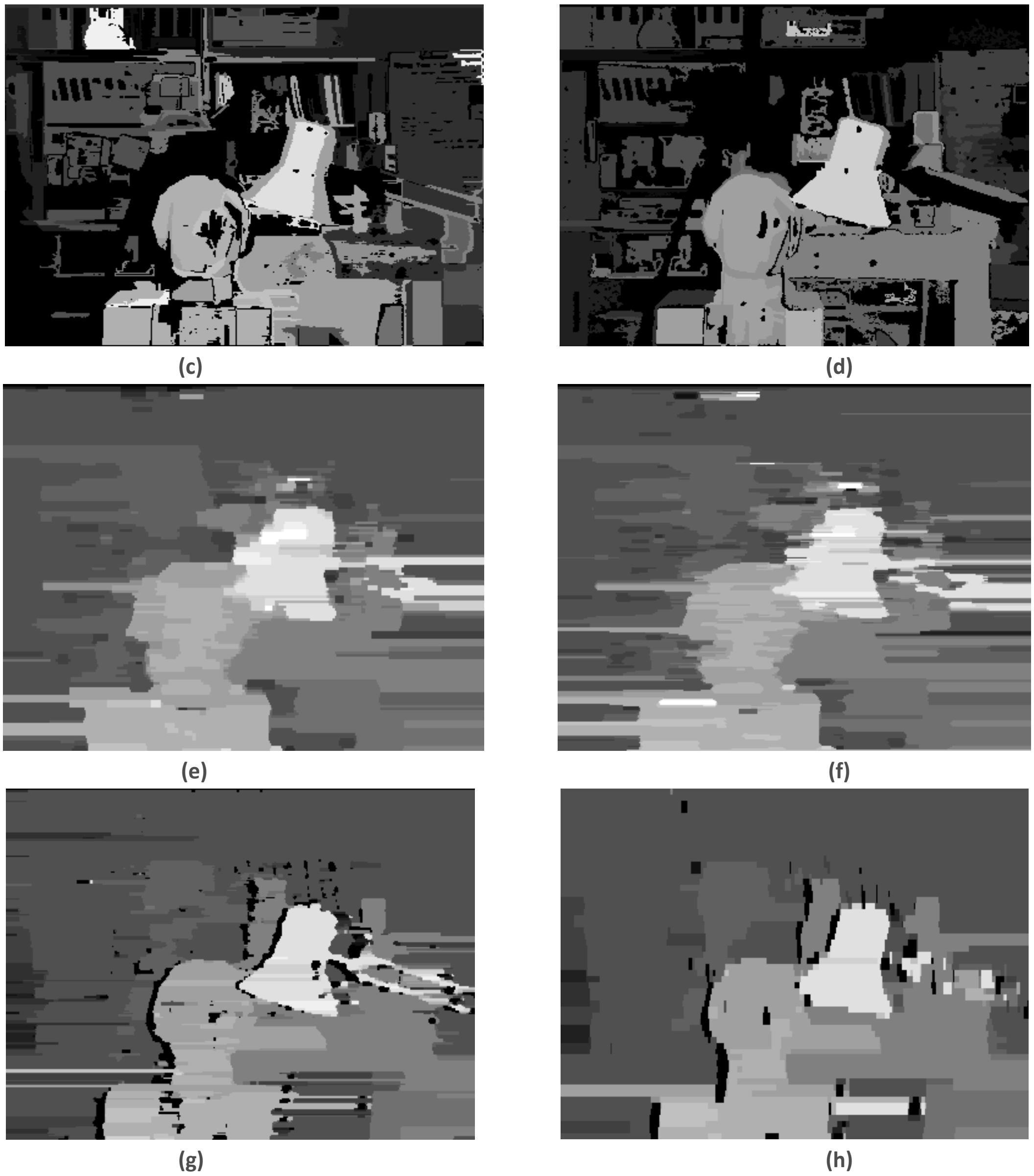
- Posibilidad de bajo coste: Un sistema de visión estéreo tan sólo necesita dos cámaras, las cuales, hoy en día, pueden conseguirse por un precio muy bajo. Detectores CMOS tridimensionales, emisores de infrarrojos o laser de la suficiente potencia, cámaras de ultrasonidos u otros dispositivos y tecnologías para la estimación de la profundidad son mucho más caros que los sensores ópticos comerciales.
- Medida absoluta de la distancia: Sistemas también muy baratos, de visión monocular, ya sea mediante análisis de estructuras o el llamado “seguimiento de puntos”, no consiguen medidas absolutas de distancia, sino tan sólo relativas, lo cual puede ser muy problemático en términos de movilidad.
- Precisión: Otros sistemas monoculares, con medida absoluta, como el de desenfoque, no consigue precisiones como los algoritmos de estereovisión.

Para ello, se fueron diseñando diversas aproximaciones, sumando un total de cuatro propuestas, hasta dar con una versión lo suficientemente efectiva y precisa. La figura 9.2 muestra una imagen de referencia, llamada Tsukuba (propuesta por al Universidad de Tsukuba), su mapa real de profundidad medido con laser (y que se toma de referencia para cálculos de error) y distintas estimaciones obtenidas con los algoritmos propuestos.



(a)

(b)



(c) (d)

(e) (f)

(g) (h)

Fig. 9.2. (a) Imagen izquierda del par Tsukuba, (b) mapa real de profundidad, (c) "gray scale region growing" (RG), (d) "pseudo-color region growing" (P-C RG), (e) "full definition points based" (FDPB), (f) "half definition points based" (HDPB), (g) "80% de líneas escaneadas con fast&dense" (80% F&D) y (h) "20% de líneas escaneadas con fast&dense" (20% F&D).

La tabla 9.3 muestra los principales resultados de los análisis de error y tiempo (sobre un PC con procesador simple de 1.6GHz).

Algoritmo	Error en los píxeles no ocultos	Tiempo de procesado (fps)
RG	55.9%	50ms (20fps)
P-C RG	46.9%	77.4ms (12fps)
FDPB	11.3%	23.2ms (43fps)
HFPB	10.5%	11.9ms (84fps)
20% F&D	8.99%	11.3ms (88fps)
80% F&D	9.98%	2.9ms (347fps)

Tabla 9.3. Resumen de los resultados de los cuatro algoritmos propuestos (y algunas variantes sobre el número de líneas escaneadas).

Adicionalmente, el algoritmo de F&D (que puede procesar cualquier porcentaje de líneas, y a partir de ahora trabajará o bien con la definición completa –FD- o media definición –HD-) sufrió unas modificaciones de gestión de memoria, modificando ligeramente la velocidad y muy sustancialmente el uso de memoria. Dicha nueva versión se llama FD ó HD F&D L (lite) (ver figura 9.3).

Podemos definir un índice de mérito para estos algoritmos como:

$$i = \frac{1}{error \cdot tiempo \cdot memoria} \qquad (9.3)$$

De dicho índice, obtenemos los siguientes resultados:

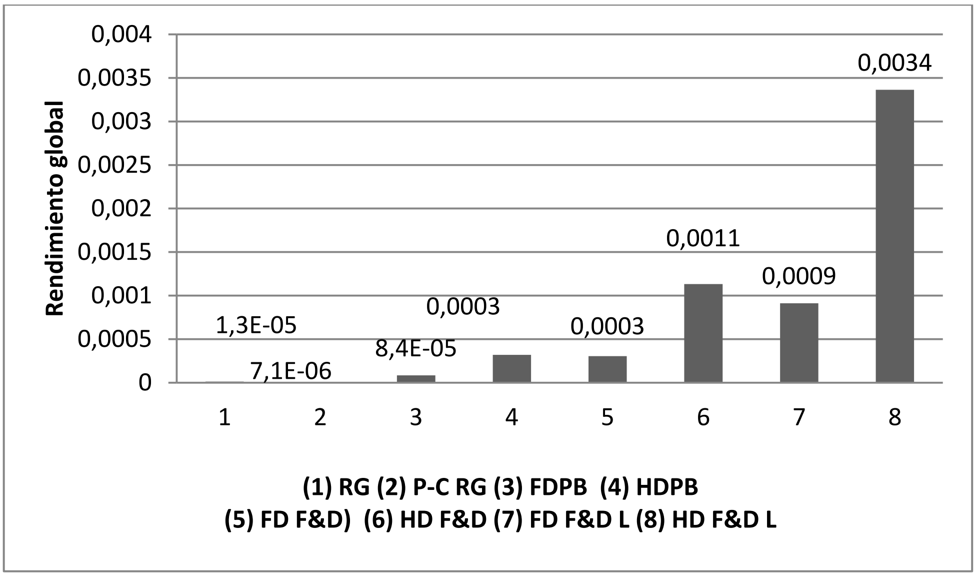


Fig. 9.3. Comparación global de los algoritmos propuestos.

El algoritmo utilizado finalmente fue el 7, F&D lite trabajando en definición total, pues demostró, en otros conjuntos de imágenes, ser más robusto frente a errores que las versiones que descartan la mitad de las líneas.

9.6. Sonificación

El capítulo sobre sonificación repasa los fundamentos de psicoacústica (la ciencia que estudia la percepción subjetiva del sonido), así como algunas de las propiedades básicas tanto del sonido como del sistema auditivo humano.

Independientemente, se realiza un estudio del estado del arte de las distintas técnicas de sonificación, es decir, de la forma en la que información no auditiva se convierte en sonido, y más específicamente sobre la traducción de imágenes a sonidos. Se puede consultar el siguiente artículo de revisión al respecto [46].

A modo de resumen, se pueden dividir las propuestas en al menos dos grandes bloques:

- Monaural: Se utiliza un solo canal para transmitir la información. Esta información suele ser codificada en una dimensión continua, tal como frecuencia, amplitud, etc. Por lo tanto, la mayoría de los sistemas que usan este paradigma son unidimensionales. Algunos ejemplos de este grupo de ETA son el OD Nottingham, Torch EE.UU., Mims, FOA Laser caña [47] o el detector Sidewalk [48].

- Binaural: La binauralidad permite transmitir información mucho más rica para el usuario. De hecho, una dimensión más. Por lo tanto, algunas de las propuestas más complejas se pueden encontrar en esta familia. Algunos de los más importantes dispositivos de esta familia son el Navbelt [27], el Multi and cross-modal ETA [49, 50], el Echolocation [51], el EAV [43], el 3-D Support [52-54], el CASBliP [29, 55], el vOICe [44], el Sonic Mobility Aid [56], el NAVI [57], el Sonic Pathfinder [58-60, 60, 61], el Sonic Guide [62, 63], etc.

En cuanto a cómo se correlacionan características espaciales y sonidos, encontramos dos grandes paradigmas:

- Psicoacústico: Este paradigma emplea la discriminación natural de la fuente según sus parámetros espaciales (distancia, azimut y altitud por ejemplo). Se utilizan funciones y curvas complejas (llamadas, en su versión más elaborada, HRTF o Head Related Transfer Functions), para simular, por medio de convoluciones, la posición virtual de la fuente.
- Arbitrario: Para los atributos que no están directamente relacionadas con los sonidos (como el color o la textura, por ejemplo), o para sustituir los parámetros espaciales en los que el error de estimación sea muy alto, muchas AT utilizan transformaciones arbitrarias entre una propiedad espacial o física y el sonido asociado, necesitando un entrenamiento.

Por supuesto, se pueden encontrar infinidad de combinaciones de estos dos paradigmas. Como resumen de dichas combinaciones, a veces dominantemente arbitrarias, otras con una carga importante psicoacústica, sirva la siguiente clasificación:

- Fish [64] utiliza la frecuencia para cartografiar la posición vertical, y la diferencia de sonoridad binaural para la posición horizontal (podemos ver aquí una transformación psicoacústica de esta segunda dimensión). El brillo se asigna al volúmen. No se debate en este momento si este brillo está relacionado con el brillo natural de la escena, o con otro parámetro tal como la profundidad, después de un procesamiento de la escena. Estos tres parámetros cambian en cada cuadro y normalmente se llama *point mapping*.
- Dallas [65] asigna de nuevo posición vertical a la frecuencia, la posición horizontal al tiempo y el brillo al volúmen. El mapeo utilizado es un ejemplo del *piano transform* (se muestra en la figura 6.27). Un ejemplo moderno de esta transformación se encuentra en el proyecto vOICe [44]. En este caso, un clic indica el comienzo de la descripción de la escena.
- Podemos agregar a éstas dos asignaciones otra que relaciona distancia con frecuencia, propuesta por Milios [66]: "Asignando a las frecuencias más altas la cercanía, haciendo hincapié de este modo en la importancia de los objetos cercanos. La frecuencia máxima en este mapeo (4200 Hz) corresponde a una medición de alcance de 0,30 m, mientras que la frecuencia mínima (106,46 Hz) corresponde a una medición de alcance de 15 m "(p. 417). Vamos a llamar a esta opción *pitch transform*. El principal problema de este enfoque es el proceso de aprendizaje para una dimensión bastante intuitiva,

como la distancia, teniendo sin embargo ventaja en el caso de entornos muy ruidosos, donde las diferencias de sonoridad puede no ser percibidas mientras que las diferencias de tono, a volumen constante, trabajan mucho mejor.

- Por último, y forzando el concepto de mapeado, se propondrá la transformada verbal, es decir, traducir la escena en voz (sintética o grabada), que permita al usuario formar una representación mental de los alrededores. Por ejemplo, el Mini-radar [67]produce mensajes como "Stop" o "Vía libre", dependiendo del entorno. La principal ventaja es que no se necesita capacitación en este caso (solo, en su caso, la lengua), desechando, por otro lado, una gran cantidad de información, pues no todas las situaciones tienen correlación con una palabra registrada en el sistema. Así, las escenas complejas son difícilmente convertibles automáticamente al habla.

Se propone, en este capítulo, una modificación del *point mapping* para resolver el problema de sonificación, siguiendo el esquema de la figura 9.4.

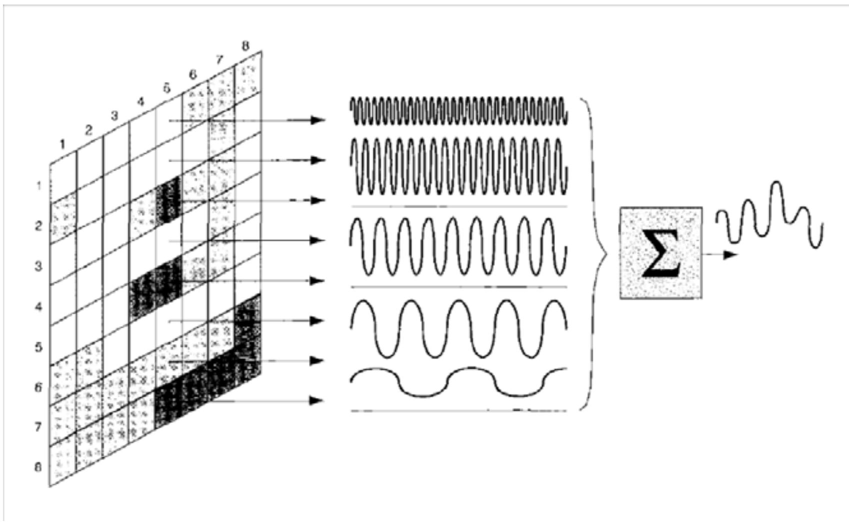


Fig. 9.4. Point mapping extraído de [68].

Los parámetros principales de dicha propuesta quedan resumidos en la siguiente lista:

- La intensidad representará inversamente la distancia (a más claridad, más cercanía, como en las imágenes mostradas en la figura 9.2). El brillo (la profundidad) se correlaciona con el volumen, pero el rango de valores posibles está discretamente dividido en 6 diferentes sonidos (voz sintética, flauta, oboe, trombón y trompeta con sordina), llegando a ser más agudo cuando los puntos se acercan más a ayudar en la distancia discriminación. En una segunda versión (para trabajar en tiempo real con cambios suaves, cosa que no permite la división en instrumentos), se utiliza un filtro paso bajo y una base estridente (trompeta), de forma que los sonidos lejanos no presentan prácticamente armónicos (ni, por tanto, estridencia), mientras que los más cercanos, siguiendo las recomendaciones de los requisitos de usuario, tienen mucha más estridencia, alertando del peligro que representan.
- La lateralización se lleva a cabo por las diferencias en la intensidad y el tiempo de cada sonido, como se describe en estudios psicoacústicos como [69]. Para evitar ambigüedades, un trémolo se aplica a las zonas laterales, teniendo en cuenta que cuanto más cerca de un lateral está un punto, tanto más profundo es el trémolo.

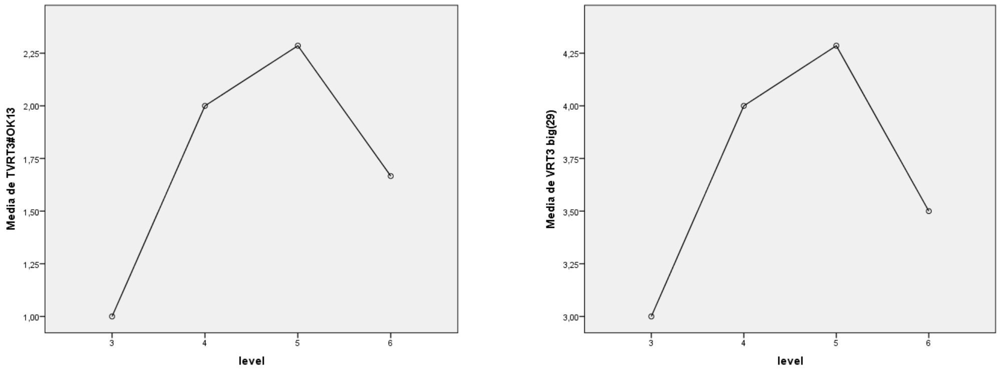
- Sólo los pixeles más cercanos (siendo su valor luminoso más alto que 42, en una gama de [0.255]) se sonificarán.
- El eje vertical está codificado por medio de notas musicales armónicas (que forman el acorde CMaj7m cuando todos los niveles de altura están excitados). Sin embargo, también se han propuesto algunos perfiles sencillos, siendo este último el más complejo. En el nivel máximo se utilizan 16 notas para la codificación altura (el acorde CMaj7m en 4 octavas), que serán reducidas a 4 en la versión hardware (HW) final). Cualquier acorde armónico permite al usuario percibir la música, en lugar de ruido desagradable.
- Se programaron 7 perfiles (numerados de 0 a 6), que cubren distintas complejidades desde una alarma booleana hasta el ya citado nivel de 16 líneas verticales y 8 horizontales.

La sonificación se lleva a cabo por medio del protocolo MIDI estándar [70], y luego mediante la variante ampliada GM2 [70].

En la sección 6.4.6 se pueden ver y oír distintos ejemplos de cada uno de los niveles.

Dicho protocolo de sonificación fue validado en las instalaciones de la facultad de psicología del Georgia Institute of Technology (GeorgiaTech) en el verano de 2012, con 28 participantes, de las cuales 17 eran mujeres y 11 varones, con una edad media de 33,46 años (entre 18 y 62). De entre ellos, 13 eran videntes, 10 tenían baja visión y 5 eran completamente ciegos. Todos oían sin mayores problemas.

Las pruebas fueron llevadas a cabo sobre entornos de realidad virtual. Dichos tests demostraron, entre otras cosas, que el nivel 6 no era funcional, tal y como demuestra la figura 9.5.



(a) (b)

Fig. 9.5. (a) Número de obstáculos correctamente detectados en función del nivel en un test de realidad virtual y (b) valoración de la facilidad para percibir obstáculos grandes (en rango de 1 a 5, escala de Likert) frente a nivel.

Otras conclusiones de este estudio están relacionadas con la necesidad de un nivel educativo adecuado para poder utilizar correctamente el dispositivo, así como los problemas que encuentran personas de mayor edad. Por último, resultó bastante claro que el uso habitual de ordenadores facilita el aprendizaje del uso de este tipo de dispositivos.

En todo caso, quedó demostrada la capacidad de los sistemas de sonificación para transmitir información del mundo espacial/visual, aunque dicha capacidad depende fuertemente de varios parámetros, cuya discusión está detallada en la sección 6.7.

9.7. Integración Software y Hardware

Las diferentes partes desarrolladas hasta este capítulo responden a las interfaces propuestas en el capítulo 4. En este capítulo se presenta la arquitectura del sistema, así como las interfaces entre los diferentes subsistemas con el fin de obtener un sistema funcional integrado.

Se realiza la integración en dos modalidades: software (SW) y hardware (HW), sirviendo la primera de las cuales para realizar una validación intermedia antes de dar el paso a la construcción de un sistema portable, barato y útil. Por tanto, ambos sistemas fueron probados con usuarios, tal y como se detallará a continuación.

9.7.1. Sistema SW

En la integración de software, se han utilizado como HW complementario dos webcams USB de bajo coste [71] con una resolución de 320 × 240 píxeles a 30 fps, y alrededor de 90º de campo de visión. Estas cámaras están unidas a un ordenador portátil. El entorno de programación es el Microsoft Visual Studio 2005, con la biblioteca OpenCV para el procesamiento de imágenes, gratuita y de código abierto. Todos los programas están escritos en ANSI C. Por último, la sonificación, como se ha dicho, fue escrita en formato MIDI, más específicamente, GM2. La figura 7.1 en el capítulo 7 muestra el diagrama de flujo del sistema SW.

La velocidad de procesado del sistema de imágenes a 320×240 y sonificando paralelamente, corriendo en un procesador de un solo núcleo a 1,6 GHz y 1 GB de memoria RAM, es de alrededor de 32fps (rango de 20 a 38) medida sobre 15 imágenes en condiciones de iluminación diferentes.

Todo el sistema en su versión de software fue ensamblado y validado en el laboratorio de sonificación de la Facultad de Psicología de la GeorgiaTech, Atlanta, GA, EE.UU., durante el verano de 2012, y con la colaboración del “Center for the Visually Impaired” de Atlanta bajo la supervisión del profesor B. Walker. La demografía de los y las participantes de este experimento es idéntica a la expuesta en el caso de la sonificación, en el capítulo anterior. Los detalles se encuentran expuestos en la sección A.II.2. En cuanto a los experimentos con 4 expertas y expertos, también en el GeorgiaTech, se pueden encontrar sus detalles en la sección A.III.2.

En el test de la mesa, en el cual las personas, vendadas en caso de tener visión normal o baja visión, las y los participantes debían encontrar la situación de diversas combinaciones de objetos en una plantilla de 3×3, usando exclusivamente los sonidos, y con las cámaras enganchadas a un casco para que la escena sonificada se correspondiera con la orientación de la cabeza. La figura 9.6 muestra el porcentaje de errores en cada una de las 9 casillas en las que estaba dividida la mesa. En la tabla 9.4 se dan los detalles de dicha figura.

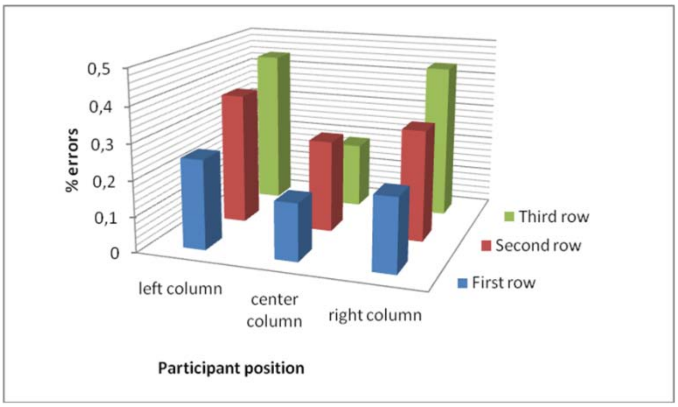


Fig. 9.6. Errores en el test de la mesa, distribuidos espacialmente según la plantilla de la mesa.

	Columna izquierda	Centro	Columna derecha
Tercera fila	.44	.18	.44
Segunda fila	.37	.26	.31
Primera fila	.25	.16	.21

Tabla 9.4. Errores medios en cada celda.

Por su parte, los y las 4 expertas, siguiendo un entrenamiento más largo y variado, estuvieron andando en otro test complementario en una habitación con objetos colocados arbitrariamente y desconocidos por ellas y ellos, debiendo encontrarlos o, al menos, no chocar con ellos.

La figura 9.7 muestra el número total de aciertos y fallos en la detección de los distintos objetos, para el total de participantes en esta prueba.

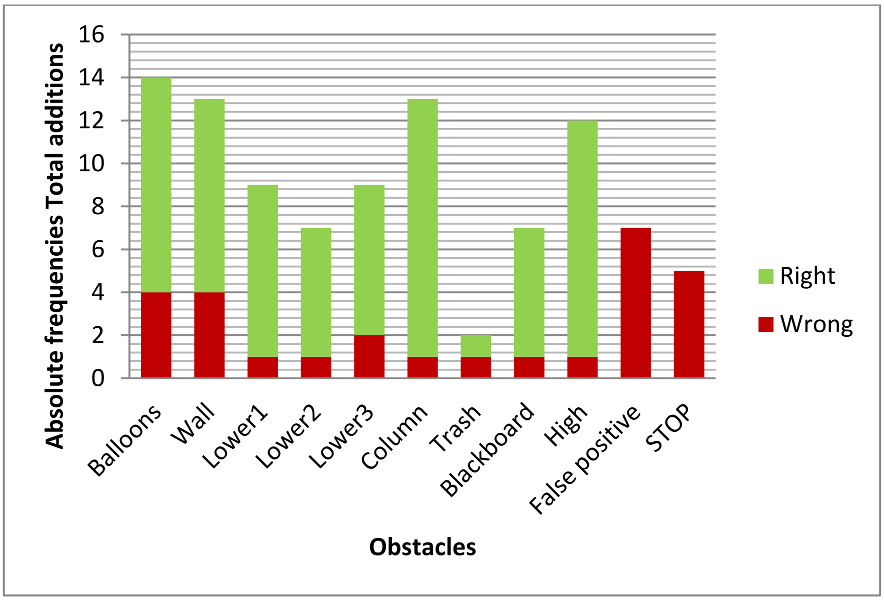


Fig. 9.7. Número total de detecciones correctas (en verde) y fallos de detección (en rojo) para cada obstáculo. Los falsos positives y las llamadas a parar (STOP) por parte del experimentador para evitar un accidente también aparecen en rojo.

Un último test para las y los expertos fue estimar la pose de una persona delante de ellas, obviamente con los ojos vendados. La figura 9.8 muestra el resultado obtenido.

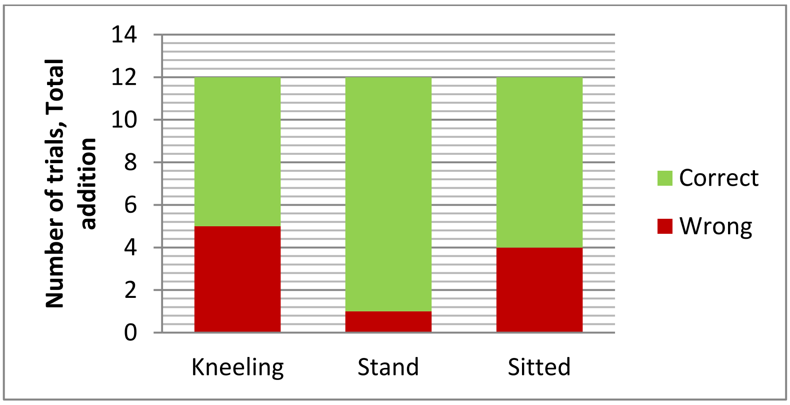


Fig. 9.8.Suma total de detecciones correctas (en verde) y erróneas (en rojo) .

Una vez más, todas las personas participantes rellenaron diversas encuestas sobre percepción subjetiva de distintos aspectos del experimento y del sistema. Nuevamente, el nivel 6, de mayor complejidad, volvió a mostrar desventajas frente al anterior que, a su vez, era el que mejor relación complejidad usabilidad pareció presentar, tal y como muestra, por ejemplo, la figura 7.9.

Las personas totalmente invidentes percibieron el sistema como menos confortable en cuanto al uso, tal y como se había encontrado en los tests de sonificación (ver figura 7.10, por ejemplo).

Según los y las expertas, el sistema funciona como visión artificial (respuesta de 4.25 en una escala Likert 5), aseguran "ver" los objetos (4.75 en la misma escala) y su opinión había mejorado tras devenir expertas y expertos (4.5), aunque los objetos pequeños aún resultan difíciles de detectar (4.25 de concordancia con esta afirmación).

Por último, se realizó una reunión (focus group) para que debatieran libremente sobre distintos aspectos del sistema, y cuyas conclusiones se detallan al final de la sección 7.1.3.2.

9.7.2. Sistema HW

Primeramente, se realizó una comparativa de tecnologías disponibles, todas ellas de coste moderado o bajo, y se compararon en la sección 7.2.2. El resumen de sus características principales se muestra en la tabla 7.4.

Se realizaron pruebas sobre microprocesadores de bajo coste (tal como el 8052), sobre plataformas Android (HTC EVO 3D) y sobre el micro-ordenador Raspberry Pi (RPi), optándose finalmente por esta última opción para realizar el sistema real, que queda explicado en la sección 7.2.3 y en la figura 7.17. Dicho sistema final utiliza las dos mismas cámaras ya presentadas, el lenguaje de programación multimedia Puredata [72] para la generación de sonidos, y un sistema de transmisión ósea para hacérselos llegar al usuario o usuaria. Complementariamente, el sistema integra una batería, un teclado numérico para seleccionar el nivel cognitivo y un amplificador de audio para ayudar a percibir los sonidos en entornos ruidosos. El nivel 6 fue eliminado tras la evaluación del sistema SW. Además, para simplificar los cálculos y mejorar el rendimiento, las imágenes finalmente utilizadas fueron de 170×120.

Dicho sistema funciona a 10.1fps con carga de trabajo completa, algo por debajo de las condiciones de tiempo real. Tiene una autonomía en la actualidad de 10h y cuesta, con elementos comprados a nivel de consumidor, 236€. Para más detalles sobre la evaluación cuantitativa, acúdase a la sección 7.2.4.1.

Finalmente, el sistema fue evaluado en condiciones reales (dentro y fuera de casa) por 8 personas de entre 22 y 60 años (edad media 41.38) en febrero de 2013 en Madrid y Las Rozas.

El nivel medio de confort elegido fue de 4.37, cercano al máximo disponible.

La transmisión ósea pareció respetar de forma casi unánime y absoluta los sonidos del mundo real (4.9 sobre una escala Likert de 5 niveles).

Sin embargo, el sistema, debido a errores en la detección, no es percibido como totalmente seguro (2.9 en la misma escala), y aunque hay obstáculos muy fáciles de detectar (como paredes, con un valor de 4.13 sobre 5), otros resultan muy difíciles (como rejas o barras, 2 y 2.43 respectivamente). Se puede consultar la tabla 7.6 para ver en detalle la facilidad de detección de diversos obstáculos.

Las principales quejas fueron, por orden de mayor a menor repetición en las diversas encuestas, la dificultad para entender la distancia, los falsos positivos del sistema, la pobre separación estéreo y la falta de control sobre el volumen. Se pueden consultar todas las respuestas ordenadas en la tabla 7.7.

Por favor, vea este video grabado tras las pruebas finales.

9.8. Conclusiones, Trabajos Futuros y Contribuciones

En el capítulo 8 se resumen las conclusiones sectoriales y globales del trabajo realizado, atendiendo a cada objetivo o hipótesis específica, y comprobando su validación o refutación.

9.8.1. Conclusiones

La tabla 9.5 resume las condiciones propuestas para cada sub-sección del sistema, así como algunas condiciones globales, y los resultados obtenidos tanto por la versión SW como por la HW.

Área	Design goal	Result SW	Results HW
Procesado de imagen	Memoria <307.2KB	230.4KB	230.4KB*
	Tasa de procesado > 24fps	32fps	10.1fps
	Precisión > 75%	75.2%	75.2%
Sonificación	Diseño para tiempo real	Sí**	Sí**
	Tasa de procesado > 24fps	44191fps	28.8fps
	Diseño basado en niveles	6 niveles	5 niveles
	Descriptores precisos	Sí (tabla 7.1)	Sí (tabla 7.1)

Sistema global	**Ancho de banda suficiente**	Sí (auriculares)	Sí (tx. ósea)
	Alarma booleana	Sí	Sí
	Diseño de base	-	Sí (fig. 7.17)
	Autonomía > 3h	-	10h
	Entrenamiento intuitivo	VR/Expertos***	Sí
	Precio < 500€	-	236€
	Mérito global > 108e-3	-	322e-3

*Supuestas imágenes de 320×240. Este sistema HW usa imágenes de 160×120 y, por tanto, la memoria utilizada es 4 veces menor.

**Con la excepción del vibrato el cual, aún así, trabaja en tiempo real.

***Entrenamiento intuitivo solo con expertos, pero no intuitivo en cuanto al resto de participantes.

Tabla 9.5. Resumen de los objetivos de diseño y resultados obtenidos.

9.8.2. Trabajos Futuros

A partir de las respuestas recopiladas de las tres evaluaciones puestas en práctica, así como de distintos problemas identificados en los resultados objetivos o en la propia percepción del autor, se han propuesto distintos trabajos futuros a realizar para solventar y/o mejorar la usabilidad, utilidad y el precio del aparato.

Dichas líneas de trabajo están divididas en grupos, según atañan al procesado de imagen, a la sonificación o al diseño HW.

En el primer caso, la principal línea de trabajo a acometer para mejorar la AT será reducir los errores del sistema de estereovisión, a fin de eliminar los falsos positivos que tanta molesta han causado a las personas participantes. Otras cuestiones a abordar serán las relativas a la selección del rango de detección y sonificación (ver figura 8.2), la división del área escaneada según modos de trabajo (búsqueda, seguir pared, global… ver figura 8.3), opciones de procesado de colores, búsqueda de objetos, OCR, etc.

En cuanto a la sonificación, se deberá mejorar la definición de los sonidos para representar más claramente la distancia, así como separar más claramente el espaciado estéreo. Además, se propuso invertir la zona de vibrato, para que los objetos centrales vibraran y los laterales no, pues esto es percibido más intuitivamente como señal de peligro que los sonidos de envolvente plana. En otro orden de cosas, aún en este campo, hay que ponderar adecuadamente el volumen total de los objetos, pues en la versión actual se da una suma lineal del volumen de cada pixel, con lo que aparecen situaciones extrañas, como que un objeto grande más lejos suene más fuerte (por tener más píxeles, aunque de menor intensidad) que un objeto cercano y más pequeño. Por último, se propuso incorporar algunos mensajes verbales sencillos, como "vía libre" o "peligro", con la posibilidad de ir completándolos y flexibilizándolos.

En cuanto al HW, habría que aumentar la potencia de cómputo, para poder incorporar algoritmos más precisos y seguir trabajando en tiempo real. Por otro lado, el control de exposición y balance de blancos de las cámaras debe estar sujeto a control, pues al ser automático, se producen desajustes como los mostrados en la figura 5.56. La autonomía, por su parte, ha resultado ser demasiado alta, necesitando una batería demasiado grande y pesada. Ésta podría ser reducida fácilmente a la mitad, aligerando el sistema total aún reduciendo la autonomía a 5h. El sistema debe incorporar un sistema de control automático de ganancia (AGC de sus siglas en inglés) y un amplificador de sonido controlable por el o la usuaria. Por último, otras cuestiones no tan críticas pero fácilmente resolubles serían incorporar distintos controles sobre el nivel deseado, unificar transmisión ósea y gafas (hasta ahora iban por separado), meter todo el sistema en una caja o incluso desarrollar un videojuego online que permita a la gente entrenarse y evaluar el sistema antes de probarlo en la vida real.

9.8.3. Contribuciones

A continuación se detallan las contribuciones a las que ha dado lugar el trabajo llevado a cabo en la investigación aquí presentada. Todas ellas pertenecen a revistas, conferencias internacionales o capítulos de libros revisados por pares y escritos en inglés.

9.8.3.1. Capítulos de Libro

- Authors: Revuelta Sanz, P., Ruiz Mezcua, B., & Sánchez Pena, J. M.
  Title: Depth Estimation. An Introduction.
  Book title: Current Advancements in Stereo Vision. 224 pp.
  Editor: Ms. Marina Kirincic. InTech Ed. In press.
  Place: Rijeka, Croatia.
  Date: 2012
  ISBN: 978-953-51-0660-9, ISBN 979-953-307-832-7
  http://www.intechopen.com/books/export/citation/EndNote/current-advancements-in-stereo-vision/depth-estimation-an-introduction

- Authors: Revuelta Sanz, P., Ruiz Mezcua, B., & Sánchez Pena, J. M.
  Title: ICTs for Orientation and Mobility for Blind People. A State of the Art.
  Book Title: ICTs for Healthcare and Social Services: Developments and Applications
  Editor: Isabel Maria Miranda & Maria Manuela Cruz-Cunha. IGI-Global.
  Place: Hershey, EE.UUDate: 2011.

9.8.3.2. Publicaciones en Revistas Indexadas en el JCR

- Authors: Pablo Revuelta, Belén Ruiz Mezcua, José Manuel Sánchez Pena, Bruce N. Walker.
  Title: Scenes and images into sounds: a taxonomy of image sonification methods for mobility applications
  Journal: Journal of the Audio Engineering Society (in press)
  Date: March 2013.

- Authors: Pablo Revuelta, Belén Ruiz Mezcua, José Manuel Sánchez Pena, Jean-Phillippe Thiran.
  Title: Segment-Based Real-Time Stereo Vision Matching using Characteristics Vectors
  Journal: Journal of Imaging Science and Technology 55(5)
  Date: Sept./Oct. 2011.
  Artículo seleccionado de especial calidad por la revista, figurando entre los 20 más descargados en Diciembre 2011.

9.8.3.3. Actas de Conferencias Internacionales

- Authors: Pablo Revuelta Sanz, Belén Ruiz Mezcua, José M. Sánchez Pena
  Title: A Sonification Proposal for Safe Travels of Blind People.
  Conference: The 18th International Conference on Auditory Display (ICAD2012).
  Publication: (not available yet).
  Place: Atlanta, Georgia (U.S.A.).
  Date: June 18-22, 2012

- Authors: Pablo Revuelta Sanz, Belén Ruiz Mezcua, José M. Sánchez Pena
  Title: Sonification as Rights Implementation.
  Conference: The 18th International Conference on Auditory Display (ICAD2012).
  Publication: (not available yet).
  Place: Atlanta, Georgia (U.S.A.).
  Date: June 18-22, 2012

- Authors: Pablo Revuelta Sanz, Belén Ruiz Mezcua, José M. Sánchez Pena
  Title: Estimating Complexity of Algorithms as a Black-Box Problem: A Normalized Time Index.
  Conference: 3rd International Multi-Conference on Complexity, Informatics and Cybernetics (IMCIC 2012).
  Publication: (not available yet).
  Place: Orlando, Florida (U.S.A.).
  Date: March 25th-28th, 2012.

- Authors: Pablo Revuelta Sanz, Belén Ruiz Mezcua, José M. Sánchez Pena
  Title: Users and Experts Regarding Orientation and Mobility Assistive Technology for the Blinds: a Sight from the Other Side.
  Conference: AIRtech 2011: Accessibility, Inclusion and Rehabilitation using Information Technologies.
  Publication: Proceedings of the AIRTech, pp. 3-4.
  Place: La Habana (Cuba).
  Date: December 13-15, 2011
  Seleccionado para ser enviado a la Journal of Research and Practice in Information Technology.

- Authors: Pablo Revuelta Sanz, Belén Ruiz Mezcua, José M. Sánchez Pena
  Title: A Review of Orientation Technologies for the Blinds
  Conference: AIRtech 2011: Accessibility, Inclusion and Rehabilitation using Information Technologies.
  Publication: Proceedings of the AIRTech, pp. 7-8.
  Place: La Habana (Cuba).
  Date: December 13-15, 2011
  Seleccionado para ser enviado a la Journal of Research and Practice in Information Technology.

- Authors: Pablo Revuelta Sanz, Belén Ruiz Mezcua, José M. Sánchez Pena
  Title: A Review of Mobility Technologies for the Blinds
  Conference: AIRtech 2011: Accessibility, Inclusion and Rehabilitation using Information Technologies.
  Publication: Proceedings of the AIRTech, pp. 5-6.
  Place: La Habana (Cuba).
  Date: December 13-15, 2011
  Seleccionado para ser enviado a la Journal of Research and Practice in Information Technology.

- Authors: Pablo Revuelta Sanz, Belén Ruiz Mezcua, José M. Sánchez Pena
  Title: Stereo Vision Matching over Single-Channel Color-Based Segmentation
  Conference: International Conference on Signal Processing and Multimedia Applications (SIGMAP 2011)
  Publication: Proceedings SIGMAP 2011, pp. 126-130.
  Place: Seville (Spain).
  Date: July, 2011.

- Authors: Pablo Revuelta Sanz, Belén Ruiz Mezcua, José M. Sánchez Pena
  Title: ATAD: una Ayuda Técnica para la Autonomía en el Desplazamiento. Presentación del Proyecto
  Conference: IV Congreso Internacional de Diseño, Redes de Investigación y Tecnología para todos (DRT4ALL 2011)
  Publication: Libro de Actas DRT4ALL 2011, pp. 151-161.
  Place: Madrid (Spain).
  Date: June, 2011

- Authors: Pablo Revuelta Sanz, Belén Ruiz Mezcua, José M. Sánchez Pena
  Title: Efficient Characteristics Vector Extraction Algorithm using Auto-seeded Region-Growing.
  Conference: 9th IEEE/ACIS International Conference on Computer and Information Science (ICIS 2010)
  Publication: Proceedings of the 9th IEEE/ACIS International Conference on Computer and Information Science (ICIS 2010), pp. 215-221.
  Place: Kaminoyama (Japan).
  Date: August, 2010.

9.8.3.4. Artículos Bajo Revisión

Por último, se han enviado los siguientes artículos, y en este momento están bajo revisión.

- Authors: Pablo Revuelta, Belén Ruiz Mezcua, José Manuel Sánchez Pena
  Title: Fast and Dense Depth Map Estimation for Stereovision Low-cost Systems
  Journal: Image Science Journal
  Date: Feb. 2013.

- Authors: Pablo Revuelta, Belén Ruiz Mezcua, José Manuel Sánchez Pena, Bruce N. Walker.
  Title: Evaluation of a Mobility Assistive Product for the Visually Impaired
  Journal: IEEE Transactions on Rehabilitation.
  Date: June. 2013.

- Authors: Pablo Revuelta, Belén Ruiz Mezcua, José Manuel Sánchez Pena, Bruce N. Walker.
  Title: Evaluation of an Artificial Vision System for the Visually Impaired
  Journal: Sensors.
  Date: May. 2013.

- Authors: Pablo Revuelta, Belén Ruiz Mezcua, José Manuel Sánchez Pena.
  Title: Implementation of a Technical Aid for an Autonomous Displacement (ATAD)
  Conference: ICAVITECH 2013.
  Date: May. 2013.

Referencias

1. AENOR, "Productos de Apoyo para personas con discapacidad. Clasificación y Terminología (ISO 9999:2007).", 2007.

2. Y. F. Heerkens, T. Bougie, and M. W. d. K. Vrankrijker, "Classification and terminology of assistive products." *International Encyclopedia of Rehabilitation*. JH Stone and M Blouin, eds. *Center for International Rehabilitation Research Information and Exchange (CIRRIE)*, 2010.

3. WHO, "Change the Definition of Blindness.", 2009.

4. WHO, "Blindness." *http://www.who.int/topics/blindness/en/index.html*, 2009.

5. WHO, "Visual impairment and blindness." *http://www.who.int/entity/mediacentre/factsheets/en/* , 2009.

6. C. Martinez, "Orientation and Mobility Training: The Way to Go." vol. Texas Deafblind Outreach. 1998.

7. H. Petrie, V. Johnson, V. Strothotte et al., "MoBIC: An aid to increase the independent mobility of blind travellers." *British Journal of Visual Impairment* vol. 15, pp. 63-66. 1997.

8. J. Reiser, "Theory and issues in research on blindness and brain plasticity." *Blindness and brain plasticity in navigation and object perception*. In: Rieser JJ, Ashmead DH, Ebner FF et al., eds. 2008. New York: Lawrence Erlbaum Associates.

9. T. Carroll, "Blindness: What it is, what it does, and how to live with it." B. Boston: Little, ed. 1961.

10. B. Lownfeld, "Effects of blindness on the cognitive functioning of children." *Nervous Child* vol. 7, pp. 45-54. 1948.

11. P. Revuelta Sanz, B. Ruiz Mezcua, and J. M. Sánchez Pena, "Sonification as a Social Right Implementation." *Proceedings of the 18th International Conference on Auditory Display (ICAD 2012)* , pp. 199-201. 2012. Atlanta, GA.

12. S. J. La Grow, "Orientation to Place." Center for International Rehabilitation Research Information and Exchange (CIRRIE), ed. vol. International Encyclopedia of Rehabilitation, pp. 1-8. 2010.

13. G. E. Hinton and L. M. Parsons, "Frames of Reference and Mental Imagery." J. Long and A. Baddeley, eds. no. 15, pp. 261-277. 1981.

14. H. Ehrlichman and J. Barrett, "Right hemispheric specialization for mental imagery: a review of the evidence." *Brain Cogn.* vol. 2 no. 1, pp. 55-76. 1983.

15. M. W. O'Boyle and J. B. Hellige, "Hemispheric asymmetry, early visual processes, and serial memory comparison." *Brain Cogn.* vol. 1 no. 2, pp. 224-243. 1982.

16. J. Hochberg, "Contemporary Theory and Research in Visual Perception." R. N. Haber, ed. vol. In the mind's eye. 1968. New York.

17. A. Noë, "Is the Visual World a Grand Illusion?" *Journal of Consciousness Studies* vol. 9 no. 5-6, pp. 1-12. 2002.

18. S. E. Palmer, *Vision Science: Photons to Phenomenology,* Cambridge,MA: MIT Press, 1999.

19. A. Ishai, "Seeing faces and objects with the "mind's eye"," *Archives Italiennes de Biologie,* vol. 148, no. 1. pp.1-9, 2010.

20. M. Behrmann, G. Winocur, and M. Moscovitch, "Dissociation Between Mental-Imagery and Object Recognition in A Brain-Damaged Patient," *Nature,* vol. 359, no. 6396. pp.636-637, 1992.

21. D. Bavelier and H. J. Neville, "Cross-modal plasticity: where and how?" *Nature Reviews Neuroscience* vol. 3 no. 443, p.452. 2002.

22. D. Yuan and R. Manduchi, "A tool for range sensing and environment discovery for the blind." *Proc.2004 Conf.Comput.Vis.Pattern Recogn.* vol. 3, p.39. 2004.

23. M. A. Hersh and M. Johnson, "Mobility: An Overview." *Assistive Technology for Visually Impaired and Blind People*. Marion A.Hersh and Michael A.Johnson, eds. no. 5, pp. 167-208. 2008.

24. R. Farcy, R. Leroux, A. Jucha et al., "Electronic Travel Aids and Electronic Orientation Aids for Blind People: Technical, Rehabilitation and Everyday Life Points of View." *Conference & Workshop on Assistive Technologies for People with Vision & Hearing Impairments Technology for Inclusion CVHI 2006*. 2006.

25. D. J. Calder, "Travel Aids For The Blind - The Digital Ecosystem Solution," *2009 7Th IEEE International Conference on Industrial Informatics, Vols 1 and 2*. pp.149-154, 2009.

26. S. Löfving, "Extending the Cane Range Using Laser Technique." *IMC9 conference Proceedings* . 2009.

27. S. Shoval, J. Borestein, and Y. Koren, "Auditory Guidance with the Navbelt-A Computerized Travel Aid for the Blind." *IEEE Transactions on Systems, Man, and Cybernetics* vol. 28 no. 3, pp. 459-467. 1998.

28. R. Audette, J. Balthazaar, C. Dunk et al., "A stereo-vision system for the visually impaired." vol. Tech. Rep. 2000-41x-1. 2000. Sch. Eng., Univ. Guelph, Guelph, ON, Canada.

29. N. Ortigosa Araque, L. Dunai, F. Rossetti et al., "Sound Map Generation for a Prototype Blind Mobility System Using Multiple Sensors." *ABLETECH 08 Conference* , p.10. 2008.

30. M. Petrella, L. Rainville, and D. Spiller, "Remote Infrared Audible Signage Pilot Program: Evaluation Report." vol. FTA-MA-26-7117-2009.01. 2009.

31. D. A. Ross, A. Lightman, and V. L. Henderson, "Cyber Crumbs: An Indoor Orientation and Wayfinding Infrastructure." *RESNA 28th International Annual Conference 2005: Atlanta, Georgia,* pp. 1-6. 2005.

32. J. Na, "The Blind Interactive Guide System Using RFID-Based Indoor Positioning System." *Computers Helping People with Special Needs, Lecture Notes in Computer Science,* vol. 4061, pp. 1298-1305. 2006.

33. J. A. Hesch and S. I. Roumeliotis, "An indoor localization aid for the visually impaired," *Proceedings of the 2007 IEEE International Conference on Robotics and Automation, Vols 1-10*. pp.3545-3551, 2007.

34. M. Sáenz and J. Sánchez, "Indoor Position and Orientation for the Blind." *HCI Part III, HCII 2009, LNCS* vol. 5616, pp. 236-245. 2009.

35. J. Sánchez and E. Maureira, "Subway Mobility Assistance Tools for Blind Users." *ERCIM UI4ALL Ws 2006, LNCS 4397* , pp. 386-404. 2007.

36. R. G. Golledge, J. M. Loomis, R. L. Klatzky et al., "Designing A Personal Guidance-System to Aid Navigation Without Sight - Progress on the Gis Component," *International Journal of Geographical Information Systems,* vol. 5, no. 4. pp.373-395, 1991.

37. HumanWare, "BrailleNote GPS." http://www.humanware.com/...gps/braillenote_gps . 2002. 1-3-2011.

38. S. K. Nagel, C. Carl, T. Kringe et al., "Beyond sensory substitution--learning the sixth sense," *J Neural Eng,* vol. 2, no. 4. pp.R13-R26, 2005.

39. A. Kooijman and M. Uyar, "Walking speed of visually impaired people with two talking electronic travel systems." *Visual Impairment Research* vol. 2 no. 2, pp. 81-93. 2000.

40. T. Kapic, "Indoor Navigation for Visually Impaired,", A project realized in collaboration with NCCR-MICS., 2003.

41. S. Treuillet, E. Royer, T. Chateau et al., "Body Mounted Vision System for Visually Impaired Outdoor and Indoor Wayfinfing Assistance." M. A. Hersh, ed. *Conference & Workshop on*

Assistive Technologies for People with Vision & Hearing Impairments Assistive Technology for All Ages, CVHI 2007. pp. 1-6. 2007.

42. ISO, "International Standard ISO 9999. Assistive products for persons with disability — Classification and terminology." . 2007.

43. J. Gonzalez-Mora, A. Rodriguez-Hernandez, E. Burunat et al., "Seeing the world by hearing: Virtual Acoustic Space (VAS)" *International Conference on Information & Communication Technologies: from Theory to Applications (IEEE Cat.No.06EX1220C)*. pp.6-ROM, 2006.

44. P. B. L. Meijer, "An Experimental System for Auditory Image Representations," *IEEE Transactions on Biomedical Engineering,* vol. 39, no. 2. pp.112-121, 1992.

45. P. Revuelta Sanz, B. Ruiz Mezcua, and J. M. Sánchez Pena, "Depth Estimation. An Introduction." *Current Advancements in Stereo Vision*. Marina Krincic, ed. 2012. InTech.

46. P. Revuelta Sanz, B. Ruiz Mezcua, J. M. Sánchez Pena et al., "Scenes and images into sounds: a taxonomy of image sonification methods for mobility applications," *Journal of the Audio Engineering Society,* vol. (in press), 2013.

47. L. W. Farmer, "Mobility devices," *Bull Prosthet Res*. pp.47-118, 1978.

48. Jie X, W. Xiaochi, and F. Zhigang, "Research and Implementation of Blind Sidewalk Detection in Portable ETA System." *International Forum on Information Technology and Applications* , pp. 431-434. 2010.

49. A. Fusiello, A. Panuccio, V. Murino et al., "A Multimodal Electronic Travel Aid Device." *Proceedings of the Fourth IEEE International Conference on Multimodal Interfaces* , pp. 39-44. 2002.

50. F. Fontana, A. Fusiello, M. Gobbi et al., "A Cross-Modal Electronic Travel Aid Device." *Mobile HCI 2002, Lecture Notes on Computer Science* vol. 2411, pp. 393-397. 2002.

51. T. Ifukube, T. Sasaki, and C. Peng, "A Blind Mobility Aid Modeled After Echolocation of Bats," *IEEE Transactions on Biomedical Engineering,* vol. 38, no. 5. pp.461-465, 1991.

52. Y. Kawai and F. Tomita, "A Support System for Visually Impaired Persons Using Acoustic Interface - Recognition of 3-D Spatial Information." *HCI,* vol. 1, pp. 203-207. 2001.

53. Y. Kawai and F. Tomita, "A Visual Support System for Visually Impaired Persons Using Acoustic Interface." *IAPR Workshop on Machine Vision App. (MVA 2000)* , pp. 379-382. 2000.

54. Y. Kawai and F. Tomita, "A Support System for Visually Impaired Persons Using Three-Dimensional Virtual Sound." *Int. Conf. Computers Helping People with Special Needs (ICCHP 2000)* , pp. 327-334. 2000.

55. M. M. Fernández Tomás, G. Peris-Fajarnés, L. Dunai et al., "Convolution application in environment sonification for Blind people." vol. VIII Jornadas de Matemática Aplicada, UPV. 2007.

56. L. H. Riley, G. M. Weil, and A. Y. Cohen, "Evaluation of the Sonic Mobility Aid." vol. American Center for Research in Blindness and Rehabilitation, pp. 125-170. 1966.

57. G. Sainarayanan, R. Nagarajan, and S. Yaacob, "Fuzzy image processing scheme for autonomous navigation of human blind." *Applied Soft Comp.* vol. 7 no. 1, pp. 257-264. 2007.

58. A. D. Heyes, "The use of musical scales to represent distance to object in an electronic travel aid for the blind." *Perceptual and Motor Skills* vol. 51 no. 2, pp. 68-75. 1981.

59. A. D. Heyes, "Human Navigation by Sound." *Physics in Tech.* vol. 14 no. 2, pp. 68-75. 1983.

60. A. D. Heyes, "The Sonic Pathfinder - A new travel aid for the blind." *In Technology aids for the disabled*. W.J.Perk and Ed. s, eds. pp. 165-171. 1983. Butterworth.

61. A. D. Heyes and G. Clarcke, "The role of training in the use of the Sonic Pathfinder." *Proceedings of the American Association for the Education and rehabilitation of the Blind and Visually Impaired, Southwest Regional Conference, Hawaii.* 1991.

62. L. Kay, "Auditory perception of objects by blind persons, using a bioacoustic high resolution air sonar," *Journal of the Acoustical Society of America,* vol. 107, no. 6. pp.3266-3275, 2000.

63. N. C. Darling, G. L. Goodrich, and J. K. Wiley, "A preliminary followup study of electronic travel aid users," *Bull Prosthet Res,* vol. 10, no. 27. pp.82-91, 1977.

64. R. M. Fish, "Audio Display for Blind," *IEEE Transactions on Biomedical Engineering,* vol. 23, no. 2. pp.144-154, 1976.

65. S. A. Dallas and A. L. Erickson, "Sound pattern generator representing matrix data format|has matrix video converted to parallel form, modulating audio tone, giving video information in terms of time and frequency." THALES RESOURCES INC, ed. 1960.

66. E. Milios, B. Kapralos, A. Kopinska et al., "Sonification of range information for 3-D space perception." *IEEE Transactions on Neural Systems and Rehabilitation Engineering* vol. 11 no. 4, pp. 416-421. 2003.

67. BESTPLUTON World Cie, "The "Mini-Radar", your small precious companion that warns you obstacles in a spoken way, and that helps you to walk straight." *http://bestpluton.free.fr/EnglishMiniRadar.htm* . Apr. 2011.

68. M. Capp and Ph. Picton, "The Optophone: An Electronic Blind Aid." *Engineering Science and education Journal* vol. 9 no. 2, pp. 137-143. 2000.

69. L. Rayleigh, "On our perception of sound direction." *Philos. Mag.,* vol. 13. pp.214-232, 1907.

70. MIDI Manufacturers Association MMA, "General MIDI 1, 2 and Lite Specifications." *http://www.midi.org/techspecs/gm.php* . 2012. 23-9-2011.

71. ICECAT, "NGS NETCam300." *http://icecat.es/p/ngs/netcam-300/webcams-8436001305400-netcam300-3943712.html* . 2013.

72. Pd-community, "Pure Data." *http://puredata.info/* . 2013.

Annex I: MIDI Messages

STATUS BYTE				DATA BYTES	
First Byte Value				**2nd Byte**	**3rd Byte**
Binary	Hex	Dec	Function		
10000000	80	128	Chan 1 Note off	Note Number	Note Velocity
10000001	81	129	Chan 2 "	(0-127)	(0-127)
10000010	82	130	Chan 3 "	see Table a1.3	"
10000011	83	131	Chan 4 "	"	"
10000100	84	132	Chan 5 "	"	"
10000101	85	133	Chan 6 "	"	"
10000110	86	134	Chan 7 "	"	"
10000111	87	135	Chan 8 "	"	"
10001000	88	136	Chan 9 "	"	"
10001001	89	137	Chan 10 "	"	"
10001010	8A	138	Chan 11 "	"	"
10001011	8B	139	Chan 12 "	"	"
10001100	8C	140	Chan 13 "	"	"
10001101	8D	141	Chan 14 "	"	"
10001110	8E	142	Chan 15 "	"	"
10001111	8F	143	Chan 16 "	"	"
10010000	90	144	Chan 1 Note on	"	"
10010001	91	145	Chan 2 "	"	"
10010010	92	146	Chan 3 "	"	"
10010011	93	147	Chan 4 "	"	"
10010100	94	148	Chan 5 "	"	"
10010101	95	149	Chan 6 "	"	"
10010110	96	150	Chan 7 "	"	"
10010111	97	151	Chan 8 "	"	"
10011000	98	152	Chan 9 "	"	"
10011001	99	153	Chan 10 "	"	"
10011010	9A	154	Chan 11 "	"	"
10011011	9B	155	Chan 12 "	"	"
10011100	9C	156	Chan 13 "	"	"
10011101	9D	157	Chan 14 "	"	"
10011110	9E	158	Chan 15 "	"	"
10011111	9F	159	Chan 16 "	"	"
10100000	A0	160	Chan 1 Polyphonic aftertouch	"	Aftertouch amount
10100001	A1	161	Chan 2 "	"	(0-127)
10100010	A2	162	Chan 3 "	"	"

10100011	A3	163	Chan 4 "	"	"
10100100	A4	164	Chan 5 "	"	"
10100101	A5	165	Chan 6 "	"	"
10100110	A6	166	Chan 7 "	"	"
10100111	A7	167	Chan 8 "	"	"
10101000	A8	168	Chan 9 "	"	"
10101001	A9	169	Chan 10 "	"	"
10101010	AA	170	Chan 11 "	"	"
10101011	AB	171	Chan 12 "	"	"
10101100	AC	172	Chan 13 "	"	"
10101101	AD	173	Chan 14 "	"	"
10101110	AE	174	Chan 15 "	"	"
10101111	AF	175	Chan 16 "	"	"
10110000	B0	176	Chan 1 Control/Mode change	See Table a1.2	See Table a1.2
10110001	B1	177	Chan 2 "	"	"
10110010	B2	178	Chan 3 "	"	"
10110011	B3	179	Chan 4 "	"	"
10110100	B4	180	Chan 5 "	"	"
10110101	B5	181	Chan 6 "	"	"
10110110	B6	182	Chan 7 "	"	"
10110111	B7	183	Chan 8 "	"	"
10111000	B8	184	Chan 9 "	"	"
10111001	B9	185	Chan 10 "	"	"
10111010	BA	186	Chan 11 "	"	"
10111011	BB	187	Chan 12 "	"	"
10111100	BC	188	Chan 13 "	"	"
10111101	BD	189	Chan 14 "	"	"
10111110	BE	190	Chan 15 "	"	"
10111111	BF	191	Chan 16 "	"	"
11000000	C0	192	Chan 1 Program change	Program #(0-127)	NONE
11000001	C1	193	Chan 2 "	See Table a1.4	"
11000010	C2	194	Chan 3 "	"	"
11000011	C3	195	Chan 4 "	"	"
11000100	C4	196	Chan 5 "	"	"
11000101	C5	197	Chan 6 "	"	"
11000110	C6	198	Chan 7 "	"	"
11000111	C7	199	Chan 8 "	"	"
11001000	C8	200	Chan 9 "	"	"
11001001	C9	201	Chan 10 "	"	"
11001010	CA	202	Chan 11 "	"	"
11001011	CB	203	Chan 12 "	"	"

11001100	CC	204	Chan 13 "	"	"
11001101	CD	205	Chan 14 "	"	"
11001110	CE	206	Chan 15 "	"	"
11001111	CF	207	Chan 16 "	"	"
11010000	D0	208	Chan 1 Channel aftertouch	Aftertouch amount	"
11010001	D1	209	Chan 2 "	(0-127)	"
11010010	D2	210	Chan 3 "	"	"
11010011	D3	211	Chan 4 "	"	"
11010100	D4	212	Chan 5 "	"	"
11010101	D5	213	Chan 6 "	"	"
11010110	D6	214	Chan 7 "	"	"
11010111	D7	215	Chan 8 "	"	"
11011000	D8	216	Chan 9 "	"	"
11011001	D9	217	Chan 10 "	"	"
11011010	DA	218	Chan 11 "	"	"
11011011	DB	219	Chan 12 "	"	"
11011100	DC	220	Chan 13 "	"	"
11011101	DD	221	Chan 14 "	"	"
11011110	DE	222	Chan 15 "	"	"
11011111	DF	223	Chan 16 "	"	"
11100000	E0	224	Chan 1 Pitch wheel control	Pitch wheel	Pitch wheel
11100001	E1	225	Chan 2 "	LSB	MSB
11100010	E2	226	Chan 3 "	(0-127)	(0-127)
11100011	E3	227	Chan 4 "	"	"
11100100	E4	228	Chan 5 "	"	"
11100101	E5	229	Chan 6 "	"	"
11100110	E6	230	Chan 7 "	"	"
11100111	E7	231	Chan 8 "	"	"
11101000	E8	232	Chan 9 "	"	"
11101001	E9	233	Chan 10 "	"	"
11101010	EA	234	Chan 11 "	"	"
11101011	EB	235	Chan 12 "	"	"
11101100	EC	236	Chan 13 "	"	"
11101101	ED	237	Chan 14 "	"	"
11101110	EE	238	Chan 15 "	"	"
11101111	EF	239	Chan 16 "	"	"
11110000	F0	240	System Exclusive		
11110001	F1	241	MIDI Time Code Qtr. Frame		
11110010	F2	242	Song Position Pointer	LSB	MSB

11110011	F3	243	Song Select (Song #)	(0-127)	NONE
11110100	F4	244	Undefined	?	?
11110101	F5	245	Undefined	?	?
11110110	F6	246	Tune request	NONE	NONE
11110111	F7	247	End of SysEx (EOX)	"	"
11111000	F8	248	Timing clock	"	"
11111001	F9	249	Undefined	"	"
11111010	FA	250	Start	"	"
11111011	FB	251	Continue	"	"
11111100	FC	252	Stop	"	"
11111101	FD	253	Undefined	"	"
11111110	FE	254	Active Sensing	"	"
11111111	FF	255	System Reset	"	"

Table A.I.1.MIDI messages specification.

2nd Byte Value			Function	3rd Byte	
Binary	Hex	Dec		Value	Use
0	0	0	Bank Select	0-127	MSB
00000001	1	1	Modulation wheel	0-127	MSB
00000010	2	2	Breath control	0-127	MSB
00000011	3	3	Undefined	0-127	MSB
00000100	4	4	Foot controller	0-127	MSB
00000101	5	5	Portamento time	0-127	MSB
00000110	6	6	Data Entry	0-127	MSB
00000111	7	7	Channel Volume (formerly Main Volume)	0-127	MSB
00001000	8	8	Balance	0-127	MSB
00001001	9	9	Undefined	0-127	MSB
00001010	0A	10	Pan	0-127	MSB
00001011	0B	11	Expression Controller	0-127	MSB
00001100	0C	12	Effect control 1	0-127	MSB
00001101	0D	13	Effect control 2	0-127	MSB
00001110	0E	14	Undefined	0-127	MSB
00001111	0F	15	Undefined	0-127	MSB
00010000	10	16	General Purpose Controller #1	0-127	MSB
00010001	11	17	General Purpose Controller #2	0-127	MSB
00010010	12	18	General Purpose Controller #3	0-127	MSB
00010011	13	19	General Purpose Controller #4	0-127	MSB
00010100	14	20	Undefined	0-127	MSB
00010101	15	21	Undefined	0-127	MSB
00010110	16	22	Undefined	0-127	MSB
00010111	17	23	Undefined	0-127	MSB
00011000	18	24	Undefined	0-127	MSB
00011001	19	25	Undefined	0-127	MSB
00011010	1A	26	Undefined	0-127	MSB

00011011	1B	27	Undefined	0-127	MSB
00011100	1C	28	Undefined	0-127	MSB
00011101	1D	29	Undefined	0-127	MSB
00011110	1E	30	Undefined	0-127	MSB
00011111	1F	31	Undefined	0-127	MSB
00100000	20	32	Bank Select	0-127	LSB
00100001	21	33	Modulation wheel	0-127	LSB
00100010	22	34	Breath control	0-127	LSB
00100011	23	35	Undefined	0-127	LSB
00100100	24	36	Foot controller	0-127	LSB
00100101	25	37	Portamento time	0-127	LSB
00100110	26	38	Data entry	0-127	LSB
00100111	27	39	Channel Volume (formerly Main Volume)	0-127	LSB
00101000	28	40	Balance	0-127	LSB
00101001	29	41	Undefined	0-127	LSB
00101010	2A	42	Pan	0-127	LSB
00101011	2B	43	Expression Controller	0-127	LSB
00101100	2C	44	Effect control 1	0-127	LSB
00101101	2D	45	Effect control 2	0-127	LSB
00101110	2E	46	Undefined	0-127	LSB
00101111	2F	47	Undefined	0-127	LSB
00110000	30	48	General Purpose Controller #1	0-127	LSB
00110001	31	49	General Purpose Controller #2	0-127	LSB
00110010	32	50	General Purpose Controller #3	0-127	LSB
00110011	33	51	General Purpose Controller #4	0-127	LSB
00110100	34	52	Undefined	0-127	LSB
00110101	35	53	Undefined	0-127	LSB
00110110	36	54	Undefined	0-127	LSB
00110111	37	55	Undefined	0-127	LSB
00111000	38	56	Undefined	0-127	LSB
00111001	39	57	Undefined	0-127	LSB
00111010	3A	58	Undefined	0-127	LSB
00111011	3B	59	Undefined	0-127	LSB
00111100	3C	60	Undefined	0-127	LSB
00111101	3D	61	Undefined	0-127	LSB
00111110	3E	62	Undefined	0-127	LSB
00111111	3F	63	Undefined	0-127	LSB
01000000	40	64	Damper pedal on/off (Sustain)	<63=off	>64=on
01000001	41	65	Portamento on/off	<63=off	>64=on
01000010	42	66	Sustenuto on/off	<63=off	>64=on
01000011	43	67	Soft pedal on/off	<63=off	>64=on
01000100	44	68	Legato Footswitch	<63=off	>64=on
01000101	45	69	Hold 2	<63=off	>64=on

01000110	46	70	Sound Controller 1 (Sound Variation)	0-127	LSB
01000111	47	71	Sound Controller 2 (Timbre)	0-127	LSB
01001000	48	72	Sound Controller 3 (Release Time)	0-127	LSB
01001001	49	73	Sound Controller 4 (Attack Time)	0-127	LSB
01001010	4A	74	Sound Controller 5 (Brightness)	0-127	LSB
01001011	4B	75	Sound Controller 6	0-127	LSB
01001100	4C	76	Sound Controller 7	0-127	LSB
01001101	4D	77	Sound Controller 8	0-127	LSB
01001110	4E	78	Sound Controller 9	0-127	LSB
01001111	4F	79	Sound Controller 10	0-127	LSB
01010000	50	80	General Purpose Controller #5	0-127	LSB
01010001	51	81	General Purpose Controller #6	0-127	LSB
01010010	52	82	General Purpose Controller #7	0-127	LSB
01010011	53	83	General Purpose Controller #8	0-127	LSB
01010100	54	84	Portamento Control	0-127	Source Note
01010101	55	85	Undefined	0-127	LSB
01010110	56	86	Undefined	0-127	LSB
01010111	57	87	Undefined	0-127	LSB
01011000	58	88	Undefined	0-127	LSB
01011001	59	89	Undefined	0-127	LSB
01011010	5A	90	Undefined	0-127	LSB
01011011	5B	91	Effects 1 Depth	0-127	LSB
01011100	5C	92	Effects 2 Depth	0-127	LSB
01011101	5D	93	Effects 3 Depth	0-127	LSB
01011110	5E	94	Effects 4 Depth	0-127	LSB
01011111	5F	95	Effects 5 Depth	0-127	LSB
01100000	60	96	Data entry +1	N/A	
01100001	61	97	Data entry -1	N/A	
01100010	62	98	Non-Registered Parameter Number LSB	0-127	LSB
01100011	63	99	Non-Registered Parameter Number MSB	0-127	MSB
01100100	64	100	Registered Parameter Number LSB	0-127	LSB
01100101	65	101	Registered Parameter Number MSB	0-127	MSB
01100110	66	102	Undefined	?	
01100111	67	103	Undefined	?	
01101000	68	104	Undefined	?	
01101001	69	105	Undefined	?	
01101010	6A	106	Undefined	?	
01101011	6B	107	Undefined	?	
01101100	6C	108	Undefined	?	
01101101	6D	109	Undefined	?	
01101110	6E	110	Undefined	?	

01101111	6F	111	Undefined	?	
01110000	70	112	Undefined	?	
01110001	71	113	Undefined	?	
01110010	72	114	Undefined	?	
01110011	73	115	Undefined	?	
01110100	74	116	Undefined	?	
01110101	75	117	Undefined	?	
01110110	76	118	Undefined	?	
01110111	77	119	Undefined	?	
01111000	78	120	All Sound Off	0	
01111001	79	121	Reset All Controllers	0	
01111010	7A	122	Local control on/off	0=off 127=on	
01111011	7B	123	All notes off	0	
01111100	7C	124	Omni mode off (+ all notes off)	0	
01111101	7D	125	Omni mode on (+ all notes off)	0	
01111110	7E	126	Poly mode on/off (+ all notes off)		
01111111	7F	127	Poly mode on (incl mono=off +all notes off)	0	

Table A.I.2. Channel control messages.

	Note Numbers											
Octave #	C	C#	D	D#	E	F	F#	G	G#	A	A#	B
-1	0	1	2	3	4	5	6	7	8	9	10	11
0	12	13	14	15	16	17	18	19	20	21	22	23
1	24	25	26	27	28	29	30	31	32	33	34	35
2	36	37	38	39	40	41	42	43	44	45	46	47
3	48	49	50	51	52	53	54	55	56	57	58	59
4	60	61	62	63	64	65	66	67	68	69	70	71
5	72	73	74	75	76	77	78	79	80	81	82	83
6	84	85	86	87	88	89	90	91	92	93	94	95
7	96	97	98	99	100	101	102	103	104	105	106	107
8	108	109	110	111	112	113	114	115	116	117	118	119
9	120	121	122	123	124	125	126	127				

Table A.I.3. Notes in Dec.

Number (Dec)	Instrument Name	Number (Dec)	Instrument Name
0	Acoustic Grand Piano	64	Soprano Sax
1	Bright Acoustic Piano	65	Alto Sax
2	Electric Grand Piano	66	Tenor Sax
3	Honky-tonk Piano	67	Baritone Sax
4	Electric Piano 1	68	Oboe
5	Electric Piano 2	69	English Horn
6	Harpsichord	70	Bassoon
7	Clavi	71	Clarinet
8	Celesta	72	Piccolo
9	Glockenspiel	73	Flute

10	Music Box	74	Recorder
11	Vibraphone	75	Pan Flute
12	Marimba	76	Blown Bottle
13	Xylophone	77	Shakuhachi
14	Tubular Bells	78	Whistle
15	Dulcimer	79	Ocarina
16	Drawbar Organ	80	Lead 1 (square)
17	Percussive Organ	81	Lead 2 (sawtooth)
18	Rock Organ	82	Lead 3 (calliope)
19	Church Organ	83	Lead 4 (chiff)
20	Reed Organ	84	Lead 5 (charang)
21	Accordion	85	Lead 6 (voice)
22	Harmonica	86	Lead 7 (fifths)
23	Tango Accordion	87	Lead 8 (bass + lead)
24	Acoustic Guitar (nylon)	88	Pad 1 (new age)
25	Acoustic Guitar (steel)	89	Pad 2 (warm)
26	Electric Guitar (jazz)	90	Pad 3 (polysynth)
27	Electric Guitar (clean)	91	Pad 4 (choir)
28	Electric Guitar (muted)	92	Pad 5 (bowed)
29	Overdriven Guitar	93	Pad 6 (metallic)
30	Distortion Guitar	94	Pad 7 (halo)
31	Guitar harmonics	95	Pad 8 (sweep)
32	Acoustic Bass	96	FX 1 (rain)
33	Electric Bass (finger)	97	FX 2 (soundtrack)
34	Electric Bass (pick)	98	FX 3 (crystal)
35	Fretless Bass	99	FX 4 (atmosphere)
36	Slap Bass 1	100	FX 5 (brightness)
37	Slap Bass 2	101	FX 6 (goblins)
38	Synth Bass 1	102	FX 7 (echoes)
39	Synth Bass 2	103	FX 8 (sci-fi)
40	Violin	104	Sitar
41	Viola	105	Banjo
42	Cello	106	Shamisen
43	Contrabass	107	Koto
44	Tremolo Strings	108	Kalimba
45	Pizzicato Strings	109	Bag pipe
46	Orchestral Harp	110	Fiddle
47	Timpani	111	Shanai
48	String Ensemble 1	112	Tinkle Bell
49	String Ensemble 2	113	Agogo
50	SynthStrings 1	114	Steel Drums
51	SynthStrings 2	115	Woodblock
52	Choir Aahs	116	Taiko Drum
53	Voice Oohs	117	Melodic Tom

54	Synth Voice	118	Synth Drum
55	Orchestra Hit	119	Reverse Cymbal
56	Trumpet	120	Guitar Fret Noise
57	Trombone	121	Breath Noise
58	Tuba	122	Seashore
59	Muted Trumpet	123	Bird Tweet
60	French Horn	124	Telephone Ring
61	Brass Section	125	Helicopter
62	SynthBrass 1	126	Applause
63	SynthBrass 2	127	Gunshot

Table A.I.4. General MIDI 1.0 instruments set.

Annex II. User Tests over Virtual Reality

A.II.1 Introduction

To improve the evaluation of the sonification protocol, regarding the online validation over static images, and ignoring the effects of the image processing and bone conduction, a virtual reality (VR) environment was set-up in order to provide the sonification subsystem noise-free 2.5D images.

This experiment was carried out in the facilities of the Sonification Lab, School of Psychology of the Georgia Institute of Technology, under the supervision of Prof. Bruce Walker, and complaining with the following requirements required by this Institute:

- Approval of the experimentation protocol by the Institutional Review Board (IRB http://www.compliance.gatech.edu/), as "Expedited review" dealing with human beings.
- Completion of the "Social / Behavioral Research Investigators and Key Personnel" course of the Collaborative Institutional Training Initiative (https://www.citiprogram.org/) by all the experimenters.

A.II.2 Demographics

In the VR tests, 17 undergraduates and postgraduates students from the Georgia Institute of Technology, plus 11 from the Center for the Visually Impaired (CVI) of Atlanta (Georgia) participated in this study. The sample was composed by 11 males and 17 females, with a mean age of 33.46 years, range 18-62. Among them, 13 were sighted, 10 presented low vision and 5 were completely blind. All reported normal or correct to normal hearing. The experts were four students of 22 years old, 3 female and one male. All of them were sighted and normal hearing.

The gathered factors are coded as follows:

- AGE: the age in years the day of the test.
- VI: three ordered values: sighted (1), low vision (2) and blind (3).
- EDU: four values: elementary school (1), high school (2), some college (3) and college degree or higher (4).
- COMP: five ordered values about the use of computers: never (1), rarely (2), once a week (3), once a day (4) and many times a day (5).

Initial analysis of the data gathered from the demographic questions of the survey raised the evidence of non-independency of some of the factors exposed in the previous section.

Table A.II.1 shows the correlations between the four demographic variables.

Pearson correlation	AGE	VI	EDU	COMP
AGE	1	.752** (p<.001)	-0.458* (p=0.014)	-0.708** (p<.001)
VI		1	-0.474* (p=0.011)	.637** (p<.001)
EDU			1	-0.527* (p=.004)
COMP				1

Table A.II.1. Pearson's correlation index and p-value. * marks significant correlations at a level of .05 and ** marks significant correlations at a level of .001.

These results are consequences of the specific characteristics of participants from both the GeorgiaTech and the CVI pools. The first group is composed mostly by sighted individuals, aged between 18 and 26 years old, using the computer many times a day and with some college as minimum educational level. The CVI participants are aged 51.91 years old with a range between 36 and 64 years old, with an average educational level of 2.7 and a computer use of 3.82. The GeorgiaTech participants had an average age of 21.51 years old (range between 18 and 26), educational level of 3.35 and computer use of 4.94.

Use of computer, age and visual impairment are highly correlated and, thus, the analysis will be done over the use of computer, since it is the most descriptive variable (the age is quite variable and the visual impairment is so narrow with only 3 different values).

A.II.3 Experiment

A.II.3.1 Hardware and Software Set-up

The VR world was designed by means of Unity3D game engine (http://unity3d.com/), which allows the definition of spatial environments, where the user can move around to experiment with different objects, ruled by physical laws as gravity.

The way the user can interact with this VR is by means of the InterSense InertiaCube3 head tracker (http://www.intersense.com/pages/18/11/) which can be seen in the following figure.

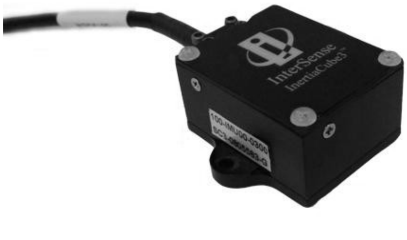


Fig. A.II.1. InertiaCube3, from Intersense.

This tracker was attached with Velcro to the headphones that the user wears, and connected through USB to the computer. It is controlled by the IServer mouse emulator program, from the same corporation.

The communication between the Unity engine and the sonification program was implemented by mean of five sequential steps:

- A TCP server is launched in the sonification program waiting for incoming connections.
- When the Unity starts rendering, it connects to the server through a TCP connection.
- The sonification program waits for a message.
- The Unity executes a Javascript routine which writes the current rendered image into a file.
- The Unity executes a C# subroutine that sends to the server an OK message {0x00, 0x00}, meaning that the file is already written.
- Whenever an OK message has been received, the image is loaded and processed. The sonification delays until the next image has been loaded and processed.

Other communications have already been explained in section 6.4.8.

The final hardware and software interconnection is shown in the figure A.II.2.

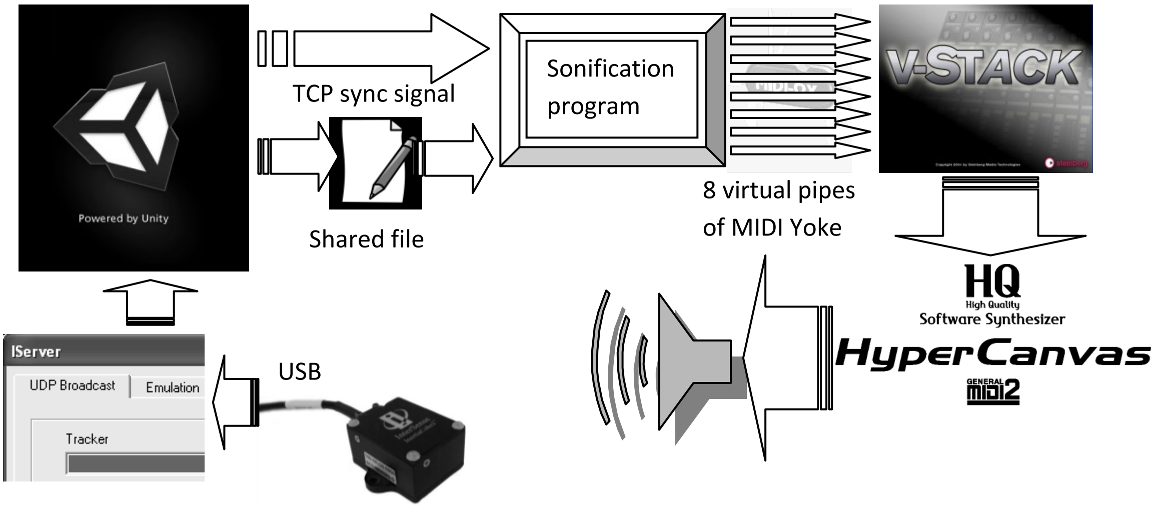


Fig. A.II.2. Hardware and Software interconnection.

A.II.3.2 VR Experiments Structure

The participants are supposed to have completed the online training for all levels. Then, they are introduced in the VR environment. However, they will only use one single level during this training, being this level assigned randomly between the 3rd and the 6th ones. The first 3 levels were not tested in VR because of their simplicity.

The VR world was structured in 5 different and independent scenes which were walked through sequentially:

- Training scene: preliminary and static scene where the participant is not blindfolded (if s/he is sighted), to become familiar with a specific sonification profile, seeing what s/he hears.
- Test 1 scene: static scene where the user, blind or blindfolded, is asked to turn around him/her self and describe what s/he hears.
- Test 2 scene: dynamic scene where the participant is asked to follow a wall at his/her right.

- Test 3 scene: dynamic scene where the participant is asked to follow a wall at his/her right, avoiding obstacles like columns or low boxes and don't getting lost because of them.
- Test 4 scene: dynamic scene where the participant is asked to explore a square room where there are different objects such as spheres, columns, boxes, etc.

These scenes will be explained more in detail in the following sections.

A.II.3.2.1. VR Training Scene

In the training scene the participant (the camera in figure A.II.3) is static, however s/he can turn around to focus on one of the 7 partial scenes, delimited as triangles in the following figure.

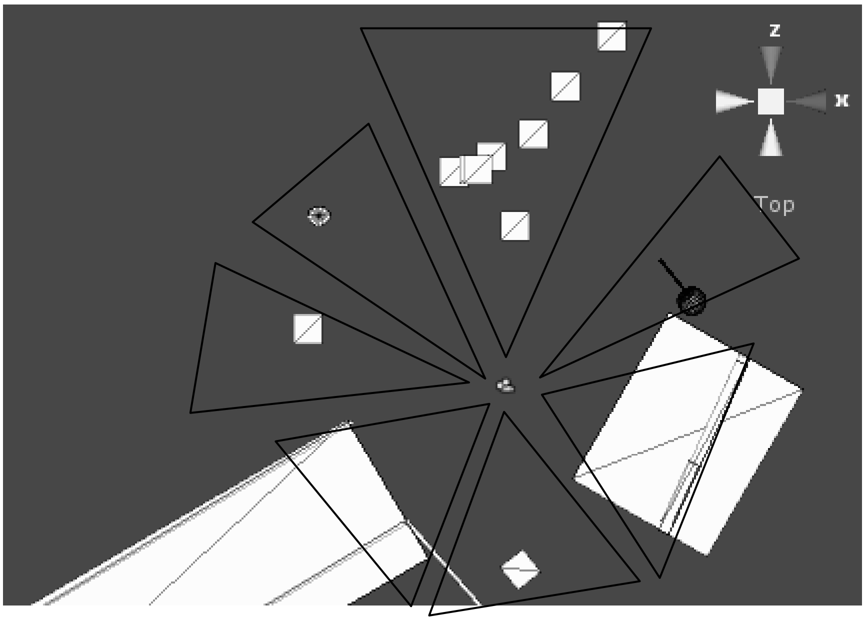


Fig. A.II.3. VR Training scene. Top view.

The scenes are composed of (from the top, clockwise direction):

- A set of boxes.
- A pendulum.
- An open door with floor.
- A box moving along a wall located at his/her right.
- A corridor.
- A box moving up and down and from farther to closer positions.
- A vertical column.

By clicking in the following link you can see a video of what the participant does, starting at the boxes (top scene of figure A.II.3). This is the 2.5D image directly processed by the sonification routine. In this link, there is a video about the same scenes and, at the bottom, the transformation following the rules of the profile 6th.

As said, the user is allowed to watch at the screen while s/he moves the head to explore the different subscenes.

A.II.3.2.2. VR Test 1

In the first test scene, the participant is blindfolded, if needed, and asked to turn around in a similar scene than that trained, but with different objects.

Now, only four subscenes are shown to the participant, whose top view is shown in the next figure.

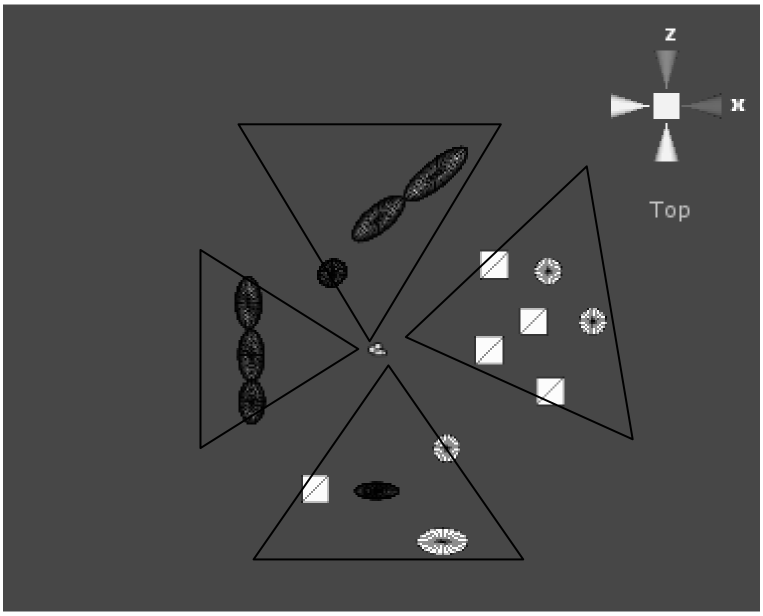


Fig. A.II.4. VR Test scene 1. Top view.

The scenes are composed of (from the left, clockwise direction):

- Three balls at different height, equal distance.
- Three balls at different height, different distance. The last one moving up and down softly.
- Six boxes and cylinders with different heights and distances.
- Four objects, one of them moving randomly horizontally.

In the following link there is a video of this scene, as perceived (if sonified) by the participant.

A.II.3.2.3. VR Test 2

In the second test scene, the participant is blindfolded, if needed, and asked to follow a wall, without getting lost.

A top view of the map is shown in the next figure.

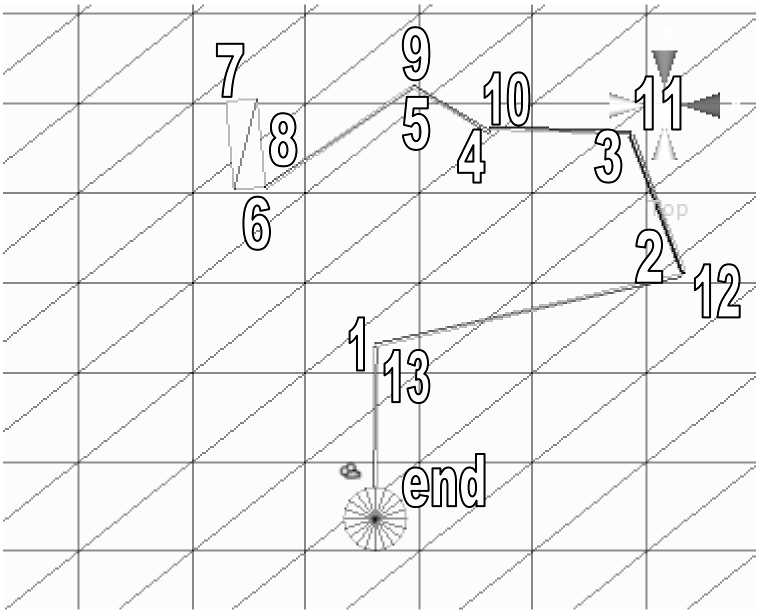


Fig. A.II.5. VR Test scene 2. Top view.

In the scene there are straight walls and corners. The user is asked to walk until s/he reaches the starting point. The corners have been labeled to identify where the users reached. There was a limit time of 10 minutes for this test.

In the following link there is a video of this scene, as perceived (if sonified) by the participant.

A.II.3.2.4. VR Test 3

In the forth test scene, the participant is blindfolded, if needed, and asked to follow a wall, without getting lost, and avoiding obstacles.

A top view of the map is shown in the next figure.

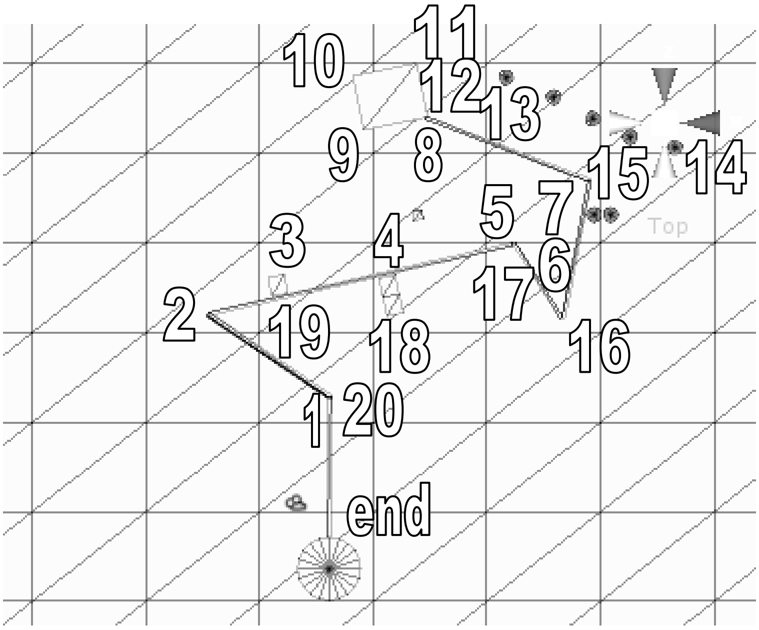


Fig. A.II.6. VR Test scene 3. Top view.

In the scene there are straight walls and corners, columns and boxes in the floor. The user is asked to walk until s/he reaches the starting point, and tell something when s/he finds an obstacle.

In the following link there is a video of this scene, as perceived (if sonified) by the participant.

A.II.3.2.5. VR Test 4

In the fifth test scene, the participant is blindfolded, if needed, and asked to explore a room, recognizing obstacles and drawing them into a sheet.

A free perspective view of the room is shown in the next figure.

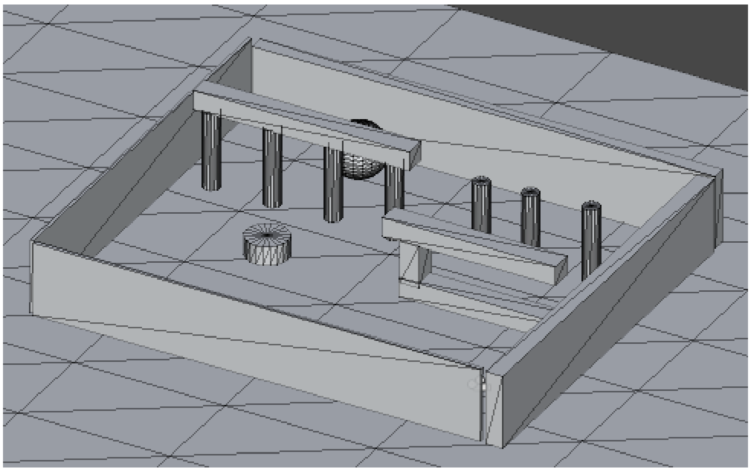


Fig. A.II.7. VR Test scene 4.

In the following link there is a video of this scene, as perceived (if sonified) by the participant.

A.II.3.3 VR Data Collection Form

The form was fulfilled by the experimenter during each experience, with the exception of the last question, where the participant was asked to draw what s/he heard moving around the virtual room. It can be accessed in this link (or after this annex in printed versions).

A.II.3.4 Survey Form

After the experiment is finished, the participant was asked to answer to a survey questionnaire about his/her experience, usability, feelings, etc. It can be accessed in this link (or after this annex in printed versions).

Data collection survey

Start Time:

End time:

Training Time:

Date: __

Participant Number: ____________________

Level chosen: _________________________

VRT1

1. Scene 1: How many objects?
2. Scene 1: Where are they located?

3. Scene 2: How many objects?
4. Scene 2: Where are they located?

5. Scene 3: How many objects?
6. Scene 3: Where are they located?

7. Scene 4: How many objects?
8. Which one is moving?
9. Scene 4: Where are they located?

Pos				
dist				

10. Scene 4: Time to finish:

Time:

Comments:

VRT2

11. Number of disorientation situation:
12. Point of arrival:

10 minutes

Comments:

VRT3

13. Number of obstacles missed:

Bloq Column Corner1 Colums LowBloq corner2

10 minutes

Comments:

15. Number of disorientation situations:
16. Did s/he perceive the columns?
17. Point of arrival:

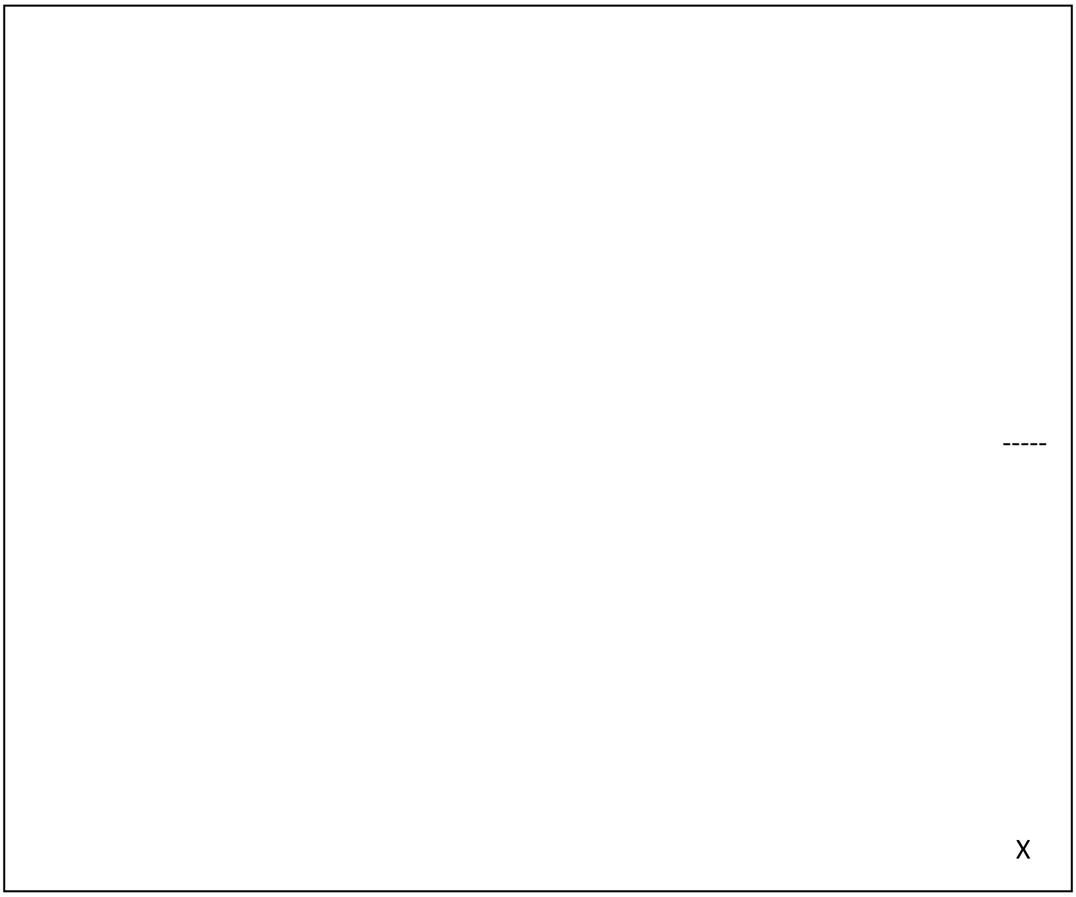
VRT4

18. Number of obstacles identified:
19. Number of disorientation situations:
20. Did s/he perceive the left columns?
21. Did s/he perceive the right columns?
22. Did s/he perceive the cylinder?
23. Did s/he perceive the sphere?
24. Did s/he perceive the horizontal obstacle?

20 minutes

Comments:

Drawing:



Date: ________________________________

Participant Number: ____________________

Level: ____

This survey asks about the training and test steps.

General Demographics

1. What is your gender? **Male** **Female**
2. What year were you born? ______________________
3. What is your level of education?

Elementary School **High School** **Some College** **College Degree or Higher**

Impairment

4. Do you have normal hearing or corrected-to-normal hearing? **Yes** or **No**
5. What would you consider to be your current level of visual impairment?

Sighted **Low Vision** **Blind**

6. In what years have you been sighted? ______________
7. In what years have you been low vision? _________________
8. In what years have you been blind? ______________________

Computer Experience

9. How many hours of experience do you have using a computer?

0 **1-10** **11-100** **More than 100**

10. About how often do you currently use a desktop computer, laptop, or note taker?

Never **Rarely** **Once a Week** **Once a Day** **Many Times a Day**

11. What relation do you have with music?

No relation **Listener** **Interpreter** **Composer** **Student**

Training step

Answer in a scale of 1-5, being 1 completely disagree and 5 completely agree with the following statements.

12. The organization of profiles makes the training easier:
13. The length of the training is long enough:

14. Write the number of the level where the training became difficult:
15. Write the number of the level in which do you feel more comfortable:

Virtual Reality Training and Test step

Answer in a scale of 1-5, being 1 completely disagree and 5 completely agree with the following statements.

Training:

16. The first scene of the VR training was easy to be understood:
17. The second scene of the VR training was easy to be understood:
18. The third scene of the VR training was easy to be understood:
19. The forth scene of the VR training was easy to be understood:

Test 1:

20. The first scene of the VR test 1 was easy to be understood:
21. The second scene of the VR test 1was easy to be understood:
22. The third scene of the VR test 1 was easy to be understood:
23. The forth scene of the VR test 1 was easy to be understood:

Test 2:

24. It was easy to follow the wall:
25. It was not tiring to follow the wall:
26. Corners present some difficulties:
27. I didn't miss the reference so often:
28. I didn't feel lost so often:

Test 3:

29. It was easy to identify the big obstacles:
30. It was easy to identify the lower obstacles:
31. Left located obstacles didn't disturb my orientation:

Test 4:

32. It was not tiring to move in the room:
33. It was easy to identify the big obstacles:
34. It was easy to identify the lower obstacles:
35. It was easy to get a mental image of the room:
36. I didn't feel lost in the room:

Reality Test step

Answer in a scale of 1-5, being 1 completely disagree and 5 completely agree with the following statements.

37. The distortions of the real system are not disturbing:
38. The system is fast enough:
39. I had to put a lot of thought into interpreting what was in front of me:
40. Please answer with the level of effort needed (1: no effort, 5 high effort):
41. I am able to imaging what I had in front of me:
42. I cannot interpret the sounds when there are more than ____ (number) objects.

General Questions

Answer in a scale of 1-5, being 1 completely disagree and 5 completely agree with the following statements.

43. The process was tiring:
44. The training should be longer:
45. The real system is more tiring than the VR one:
46. The VR step helps using the real system:
47. I'd feel safe using this system in open spaces:
48. I'd keep using dog guide or white cane (if applicable):
49. I'd buy this system if it cost less than (write a price in $):
50. Do you have any other comments?

Thank you for completing this survey!

Annex III. Experts Training and Testing

A.III.1 Introduction

To improve the evaluation of the sonification protocol, regarding the online validation over static images, and ignoring the effects of the image processing and bone conduction, a real system set of tests was designed.

This experiment was carried out in the facilities of the Sonification Lab, School of Psychology of the Georgia Institute of Technology, under the supervision of Prof. Bruce Walker, and complaining with the following requirements required by this Institute:

- Approval of the experimentation protocol by the Institutional Review Board (IRB http://www.compliance.gatech.edu/), as “Expedited review” dealing with human beings.
- Completion of the “Social / Behavioral Research Investigators and Key Personnel” course of the Collaborative Institutional Training Initiative (https://www.citiprogram.org/) by all the experimenters.

A.III.2 Demographics

In the training and testing of experts, 4 participants were enrolled with the following demographic characteristics:

- 22 years old all of them.
- Three females and one male.
- Two of them in the college, another two in a postgraduate educational level.
- All of them with normal hearing and sight.

A.III.3 Training and Experiments

A.III.3.1 Steps

The VR world was designed by means of Unity3D game engine (http://unity3d.com/), which allows the definition of spatial environments, where the user can move around to experiment with different objects, ruled by physical laws as gravity.

Step 1: Online training

10-15 min

Step 2: VR training and testing

1h30

Step 3: Table test 1	30 min
Step 4 and 5: Steps 2 and 3 repeated once	2h
Step 6: Table test 2	30 min
Step 7: Free play with objects	20 min
Step 8: Table test 3	10 min
Step 9: Walk around in a known room	25 min
Step 10: Walk around in an unknown room	30 min
Step 11: Pose estimation	10 min
Step 11: Survey	5 min
Step 12: Focus group	1h

A.III.3.1.1 Table Test

The table test is the only test of the real-system applied to every participant in the study.

The setup of the table test consisted on a computer running the stereovision algorithm used to build the depth map of the scene and the sonification program. A couple of webcams, attached to a helmet, captured the scene which was transmitted through two USB cables to be processed. The produced sound was transmitted to the user through a pair of earphones. One table of 1×1m^2 and drawn lines on it dividing it in a 3×3 grid hosted objects in different spatial combinations. The objects were a plastic glass, a spray bottle, a cover of a camera and a balloon. Figure A.III.1 shows an image of this table.

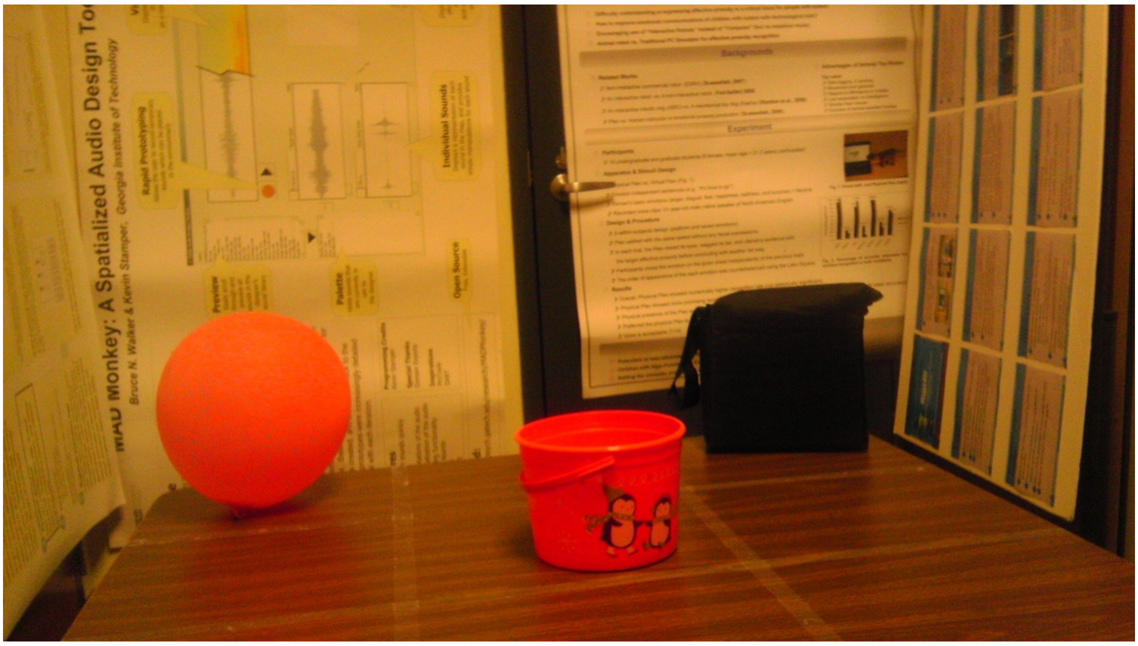

MAD Monkey: A Spatialized Audio Design To
Bruce N. Walker & Kevin Stamper, Georgia Institute of Technology
Rapid Prototyping
Individual Sounds
Preview
Palette
Open Source
Backgrounds
Experiment

Fig. A.III.1. Table with three objects.

The participant, blindfolded if needed (sighted or low vision cases), was sitting in front of the table, at 20 cm from the edge, wearing the real system, had to guess where were the objects every time the experimenter said “go ahead” to indicate that the new combination of objects was ready. They didn’t know the amount of objects on the table. Whenever they thought they had found all of them (or they were tired enough), they could say “change” to ask the experimenter to change the combination to the following one. 9 combinations of two or three objects were disposed in total. Before starting the test, a short training was done during 3 minutes, explaining the experimenter where they were some simple combinations of objects in different parts of the table. No time limit was established for this test.

Each time the user reported a position for an object, if the object was indeed in that position, a correct mark was written down. If no object was in that position, an incorrect mark was noted. If some existing object was not reported, again an incorrect mark was added to the list. Notice that this is an extremely hard way to compute the errors, since when a participant just fails in calculating the distance (but understanding there was an object in that direction), this counts as a double error, the false positive and the false negative.

The nine combinations are shown in the sheet to be fulfilled by the experimenter, in the following page.

RT1

G		F	▨			
	C		▨			
			▨			

RT2

		F	▨			
C			▨			
			▨			

RT3

G		F	▨			
			▨			
		C	▨			

RT4

G	B	F	▨			
			▨			
			▨			

RT5

G		F	▨			
			▨			
	B		▨			

RT6

G		F	▨			
	B		▨			
			▨			

RT7

			▨			
B		V	▨			
			▨			

RT8

			▨			
B	V		▨			
			▨			

Time RT1:

Time RT2:

Time RT3:

Time RT4:

Time RT5:

Time RT6:

Time RT7:

Time RT8:

Sheet to collect data from the Table test.
G: ballon
C: bucket
F: cover
V: glass

A.III.3.1.2 Free Play with Objects

During 20 minutes, the experts were left in the room with the table and objects, eyes opened, to play freely with them, manipulate, put them on the table, try different distances, etc. The experimenter was present and could help them in some configurations of objects, following their petitions, if any.

A.III.3.1.3 Known Room Step

Given that all of them were sighted, an important step of the training was done eyes opened in a room full of soft obstacles (made of paper or soft plastic) to allow them walking for 25 minutes in it.

Some of them asked the experimenter to tell them whenever they were going to crash closing the eyes, to learn to figure out the surroundings just with the sounds.

They stand in front of the different obstacles, trying to correlate the perceived sounds and the knowledge of the object seen just a moment ago.

The size of the space was 5.1×4.6 m^2.

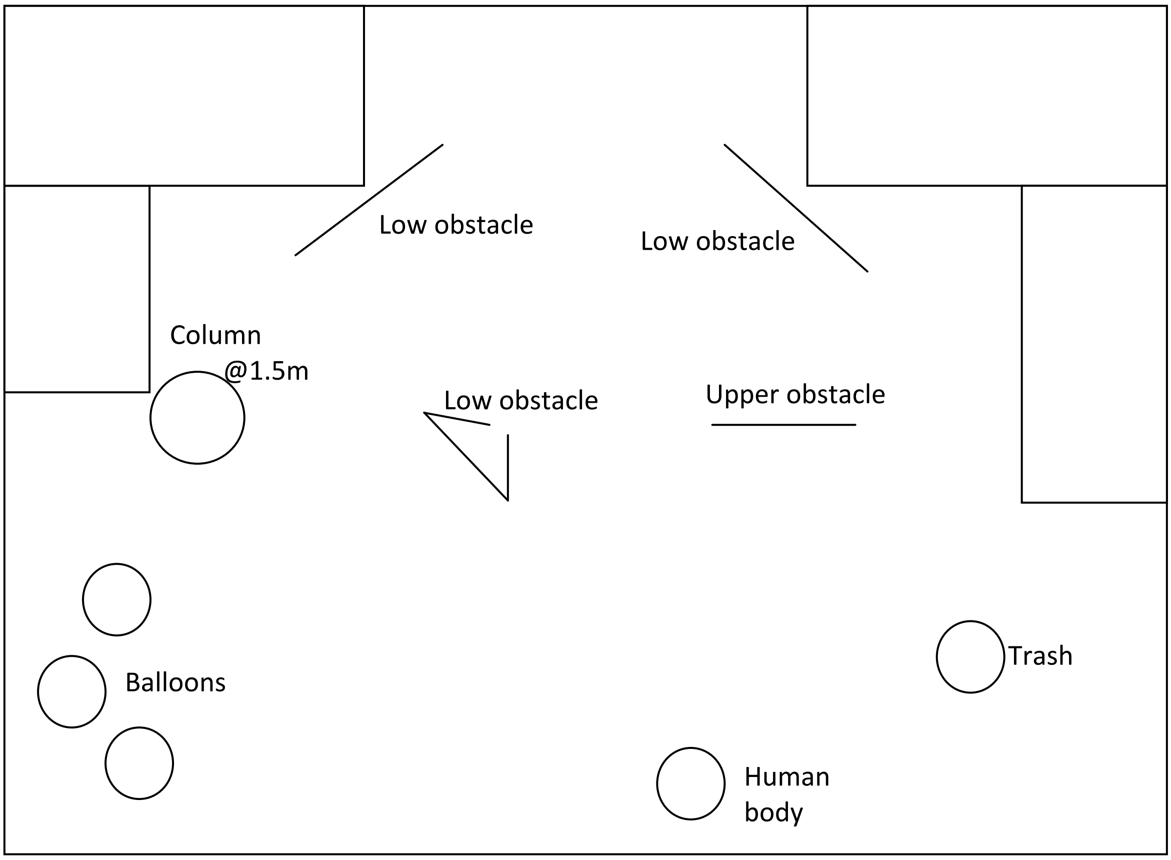


Fig. A.III.2. Map of the known room.

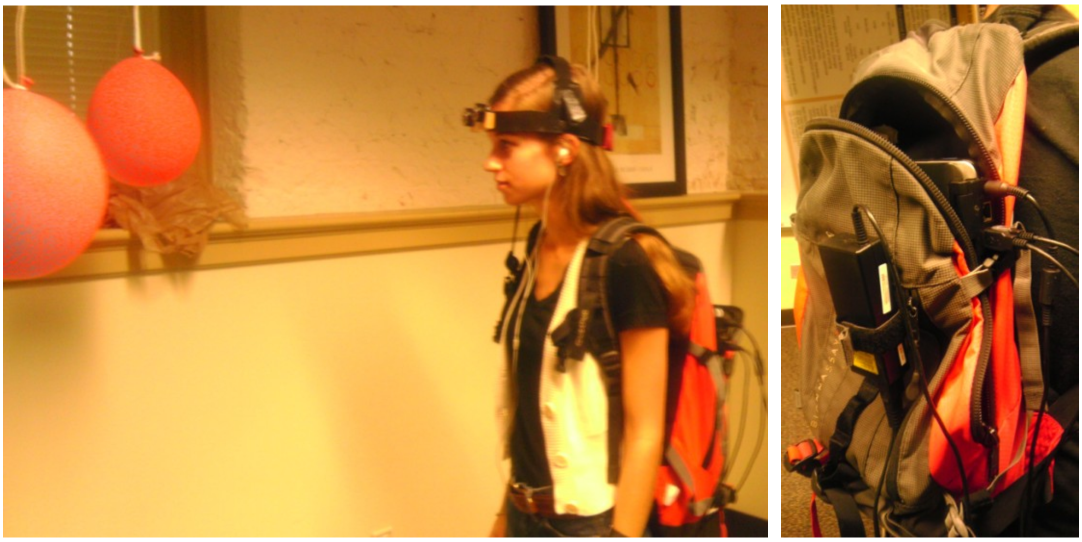


Fig. A.III.3. (left) One of the experts stands in front of the balloons, eyes opened. She wears a helmet with the webcams attached to it, the earphones and the back bag with the laptop. (Right) A detail of the back bag used to carry the laptop.

A.III.3.1.4 Unknown Room Step

The last test consists on a true mobility and artificial vision test. The objects of the room where rearranged in a different configuration and the experts were left at the "starting point" pointing to the balloons. They were asked to walk and identify each time they find an object in front of them. The experimenter wrote down all the correct and erroneous detection of obstacles or walls, as well as the number of times he had to say "STOP" to avoid a crash with some undetected obstacle.

This test lasted 30 minutes.

A video can be accessed through the following link.

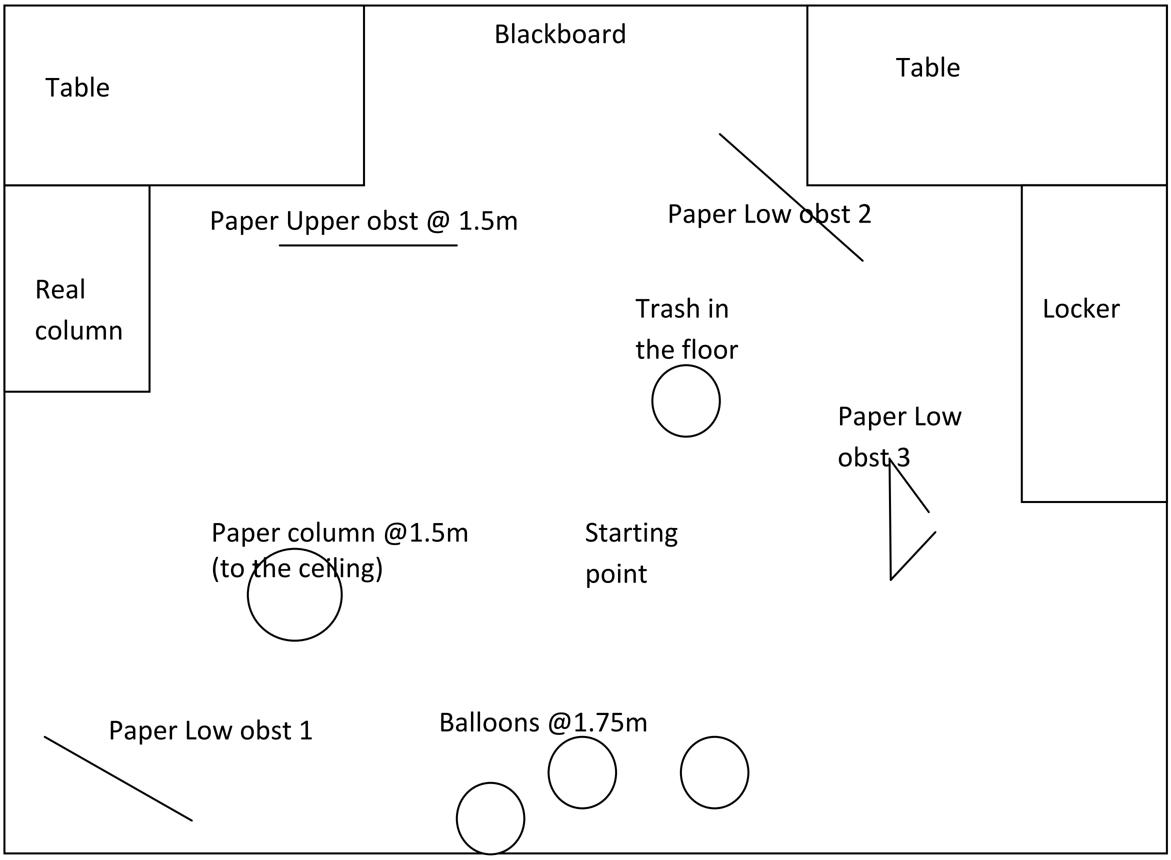


Fig. A.III.4. Map of the known room.

A.III.3.1.5 Pose Estimation

Standing up at 1m far from the experimenter, blindfolded and with the software version of the complete system, the experts were asked to guess, in 9 occasions, the pose of the experimenter, among three cases: standing up as well, sitting in a chair, and kneeling on the ground. Only correct or incorrect answers were written down.

A.III.3.1.6 Survey

They were asked to fulfill a survey at the end of the process, to evaluate changes in their appreciations about the system after this longer training and testing.

A.III.3.1.7 Focus Group

On September 19th 2012 took place a videoconference among the four experts and the author of this work for one hour where they discussed about different aspects of the system and the process.

The discussion was organized in three main areas:

- The system as a mobility aid: pros and cons, main problems and solutions suggested, etc.
- The system as an artificial system: pros and cons, main problems and solutions suggested, etc.
- General evaluation of the process, advices for the upcoming steps of the research, adaptations for visually impaired, etc.

Annex IV. User Tests over Final Hardware

A.IV.1 Introduction

After the implementation of the ATAD over the Raspberry Pi board, as explained in 7.2.3, the final system was tested with potential users in several situations, some of them of the real life.

This experiment was carried out at the homes of each volunteer, because although it produces some differences in the performance of the system (depending on the specific illumination, decoration, walls, etc. of each house), we can take advantage of the knowledge of the house every people have. The tests took place in Madrid, between February the 12nd and the 16th 2013.

A.IV.2 Demographics

In the tests participated 8 completely blind people (with at least 22 years of blindness), 4 females and 4 males with ages ranged between 22 and 60 (mean 41.38). Seven of them had normal hearing, and one of them used a hearing aid in her right ear. Three used guide dog in their displacements, the rest of them white cane.

The gathered factors are coded as follows:

- AGE: the age in years the day of the test.

- EDU: four values: elementary school (1), high school (2), some college (3) and college degree or higher (4).

- COMP: five ordered values about the use of computers: never (1), rarely (2), once a week (3), once a day (4) and many times a day (5).

No statistically significant dependency was found between the demographic variables (smallest bilateral significance found was 0.22) and, hence, we can deal with them as if they were completely independent.

Important data extracted from this test is the high familiarity with computers expressed by all the participants (the maximum one actually). Another interesting demographic datum is the educational level, with a mean of grade or university degree, and only a couple of them having Bachelor degree. Finally, five of them had some other relation with music other than listeners (we had 3 interpreters, 4 students and one composer).

A.IV.3 Experiment

A.IV.3.1 Hardware and Software Set-up

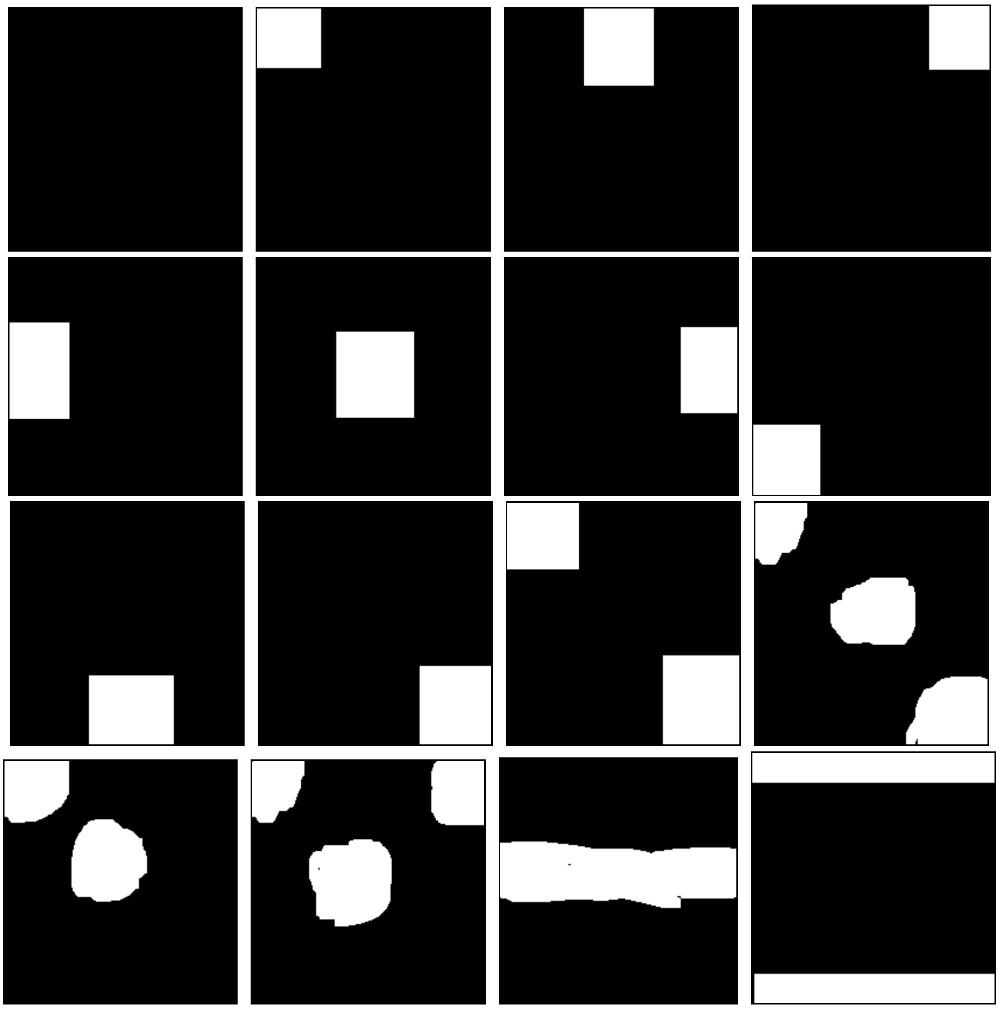
The experiment was not carried out completely on the RPi. The first session, prepared as training, was done over a virtual reality (VR) environment, as described in A.II.3.1, with the exception of the head-tracker, which was not available in this test. The scenes were moved with the mouse.

For the rest of the tests, the RPi was configured and connected as explained in 7.2.3.

A.IV.3.2 Experiments Structure

The test was divided in 4 steps:

- Training step: The participant is introduced to the system's sounds by means of static and dynamic scenes. The first ones were canonical situations of different objects, as shown in the following figure.



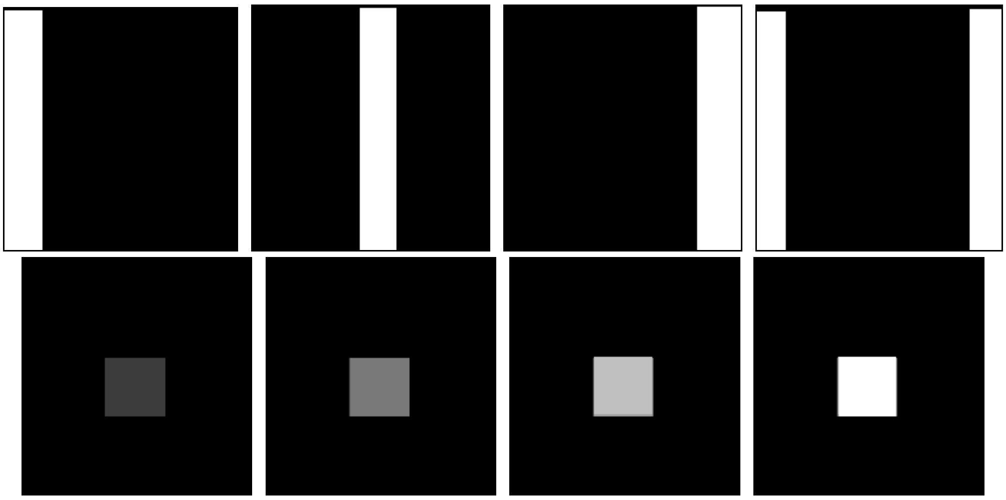


Fig. A.IV.1. Static scenes presented to the participant.

After trying all the images with all the levels (except for level 0 and some other redundancies with images and levels), the participants were asked to observe through sounds the dynamic images, in the following scene:

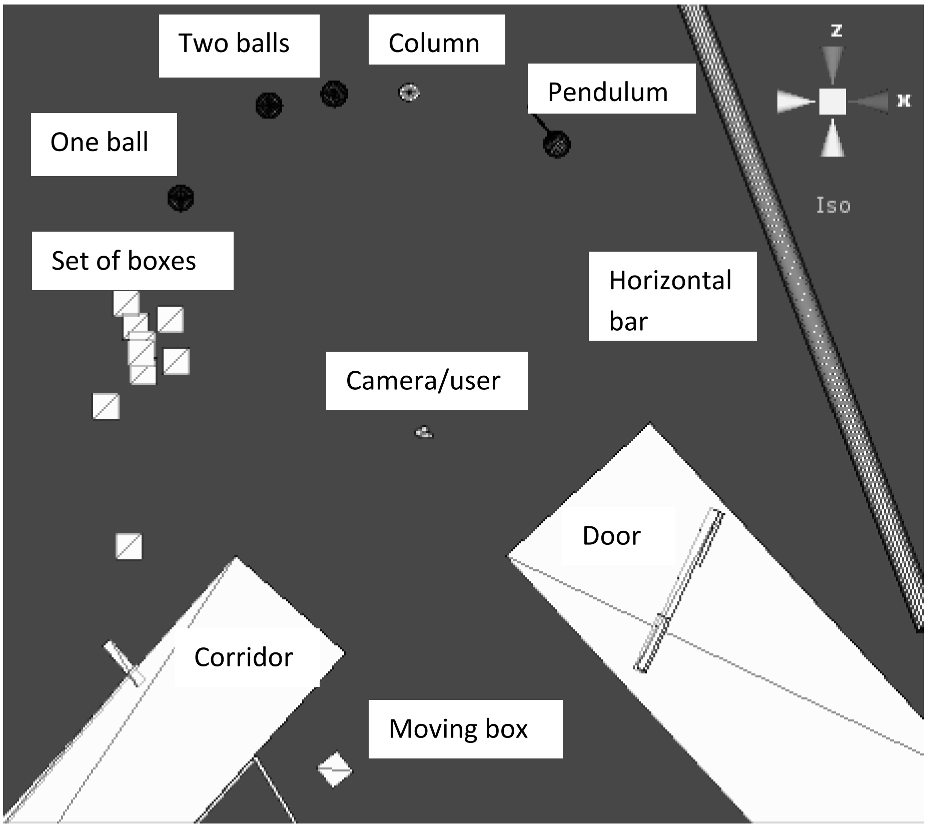


Fig. A.IV.2. VR training scene. Top view.

After this test, done with each profile, they were asked to choose one single level in which they should test the system in the following steps.

- Step 2: Objects on the table. Session in which the participant must sit in front of an empty table, and the experimenter proposes different objects in different locations so the participant must guess position and height.

- Step 3: Moving at home. In this step, the participant was asked to move freely at home, pointing to different places and spaces of their house, such as rooms, corridors, mirrors, different walls or furniture.
- Step 4: Moving outside. The final step consists of going outside, with the help of the white cane or the guide dog, and walk through usual paths. The experimenter tried them to encounter different furniture, such as rubbish containers, semaphores, tries, baskets, walls, holes, scaffolds, entrances, cars, children parks, etc.

In the following link there is a video of the first step.

In the following link there is a video of the second step.

In the following link there is a video of the third step.

In the following link there is a video of the fourth step.

A.IV.3.3 VR Data Collection Form

After each step was finished, the participant was asked to answer to a survey questionnaire about his/her experience, usability, feelings, etc. which was read for them by the experimenter. The survey can be accessed after this annex.

Nombre y número de participante: ____

Nivel: ____

Datos generales

25. ¿Género? **Femenino** **Masculino**
26. Año de nacimiento ______________________
27. Nivel educativo

    Primaria **Secundaria** **Grado/Universidad** **Postgrado**

Discapacidad

28. ¿Tienes audición normal? **Sí** **No**
29. ¿Cuál considerarías que es tu nivel actual de discapacidad visual?

    Vidente **Baja visión** **Invidente**

30. ¿Durante cuántos años fuiste vidente? ______________
31. ¿Durante cuántos años tuviste baja visión? _________________
32. ¿Durante cuántos años tuviste ceguera? ____________________

Experiencia informática

33. ¿Cuán a menudo utilizas un ordenador?

Nunca **Rara vez** **Una vez a la semana** **Una vez al día** **Muchas veces al día**

34. ¿Qué relación tienes con la música?

    Ninguna **Oyente** **Intérprete** **Compositor/a** **Estudiante**

Sesión 1

Fecha: ______________________________ Duración: ___________________________________

Responder en una escala del 1 al 5, significando 1 completamente en desacuerdo con lo afirmado, y 5 completamente de acuerdo.

35. Mi comprensión acerca de los sonidos utilizados ha mejorado a lo largo de la sesión:
36. La sesión de entrenamiento debería ser más larga:
37. Me resulta sencillo situar los objetos en el eje horizontal:
38. Me resulta sencillo situar los objetos en el eje vertical:
39. Me resulta sencillo situar los objetos según la distancia:
40. Las imágenes estáticas me resultan fáciles de entender:
41. En las imágenes estáticas me resulta difícil distinguir más de ____ objetos.
42. Las imágenes dinámicas ayudan a entrenar el sistema:
43. Escenarios complejos (pasillos, puertas…) se me hacen difíciles de entender:
44. Me resulta complicado entender los sonidos a partir del nivel:
45. El sistema de transmisión ósea (TO) me permite escuchar los sonidos reales:
46. El TO me parece una solución aceptable para transmitir sonidos:
47. El TO me molesta o incomoda:
48. El nivel en el que me siento más cómod@ es el:
49. Mi nivel de motivación para seguir con el proceso, del 1 al 5, es:
50. ¿Algún otro comentario sobre la primera sesión?

Sesión 2

Fecha: ______________________________ Duración: ______________________________

Responder en una escala del 1 al 5, significando 1 completamente en desacuerdo con lo afirmado, y 5 completamente de acuerdo.

51. Mi comprensión acerca de los sonidos utilizados ha mejorado a lo largo de la sesión:
52. El sistema real presenta más errores que el virtual:
53. Los errores del sistema real me dificultan su comprensión:
54. Los errores del sistema real me generan desconfianza:
55. El sistema real me parece demasiado lento:
56. El sistema portátil me puede resultar de utilidad:
57. El peso y la forma del sistema real es adecuado:
58. Gafas y TO son incómodas:
59. Distingo fácilmente si una persona está de pié o sentada a mi lado:
60. Me resulta sencillo entender la configuración de los objetos que hay sobre la mesa:
61. Los objetos pequeños me resultan difíciles de distinguir:
62. Muchos objetos sobre la mesa me generan confusión:
63. Mi nivel de motivación para seguir con el proceso, del 1 al 5, es:
64. ¿Algún otro comentario sobre la segunda sesión?

Sesión 3

Fecha: _______________________________ Duración: ___________________________________

Responder en una escala del 1 al 5, significando 1 completamente en desacuerdo con lo afirmado, y 5 completamente de acuerdo.

65. Mi comprensión acerca de los sonidos utilizados ha mejorado a lo largo de la sesión:
66. El sistema me parece de utilidad para uso en interiores:
67. El sistema no me da seguridad para ser usado en exteriores:
68. El sistema me ha permitido apreciar nuevos matices de mi espacio conocido:
69. El sistema me permite situarme más fácilmente en el espacio:
70. Mi nivel de motivación para seguir con el proceso, del 1 al 5, es:
71. ¿Algún otro comentario sobre la tercera sesión?

Sesión 4

Fecha: ______________________________ Duración: __________________________________

Responder en una escala del 1 al 5, significando 1 completamente en desacuerdo con lo afirmado, y 5 completamente de acuerdo.

72. Mi comprensión acerca de los sonidos utilizados ha mejorado a lo largo de la sesión:
73. El sistema me parece de utilidad para uso en exteriores:
74. Durante la sesión he ido ganando en confianza en el sistema:
75. Utilizar el sistema me da seguridad:
76. El sistema me ha permitido apreciar nuevos matices del espacio conocido:
77. Me resulta sencillo identificar:
    a. Semáforos/árboles:
    b. Personas:
    c. Buzones:
    d. Andamios:
    e. Vallas de obra:
    f. Verjas:
    g. Coches:
    h. Paredes:
    i. Portales:
    j. Perros:
    k. Huecos:
    l. Toldos:
    m. Sillas/mesas:
78. El sistema es de utilidad como ayuda a la movilidad:
79. Mi valoración general del sistema ha aumentado con el tiempo de uso:
80. Mi valoración final del sistema es, del 1 al 5, de:
81. ¿Algún otro comentario sobre la cuarta sesión?

Cuestiones generales

Responder en una escala del 1 al 5, significando 1 completamente en desacuerdo con lo afirmado, y 5 completamente de acuerdo.

82. El proceso completo fue cansado:
83. El entrenamiento debería ser más largo:
84. El sistema real es más cansado que el virtual:
85. El sistema virtual ayuda a entender y utilizar el sistema real:
86. El sistema de transmisión ósea permite recibir los sonidos del mundo real:
87. Me siento segur@ utilizando el sistema en espacios abiertos:
88. Seguiría usando bastón o perro-guía aún si dispusiera de este aparato:

¿Tienes algún otro comentario que quieras hacer?

¡¡¡Muchas gracias!!!