Seeing the world by hearing: Virtual Acoustic Space (VAS) a new space perception system for blind people.

[1]González-Mora, J.L., [1]Rodríguez-Hernández, A.F., [2]Burunat E., [1]Martin F., [2]Castellano M.A.
[1] Department of Physiology, University of La Laguna and [2]Department of Neuro-psychobiology. University of La Laguna
La Laguna, Tenerife. 38071. Spain; e-mail: jlgonzal@ull.es (José Luis González-Mora)

Abstract

Virtual Acoustic Space (VAS) is a research and development project on the perception of space using only sound. A portable electronic prototype that allows blind people to receive spatial information of their surroundings has been developed . This information is perceived via an audible image using Head Related Transfer Function (HRTFs) processed sounds. The main goal is to create for the user the illusion that the surrounding objects are covered by small sound continuously emitting sources in a particular and sustained way. Therefore, a virtual acoustic world is generated, where a physical object emit sounds from all the coordinates of its surface. Our results validate the hypothesis that, it is possible to generate an experience of global and sustained presence of different objects inside the perception field from these stimuli, with the same shape, dimensions and location as the real environment.

Our objectives are now focused on: a better delimitation of the observed capabilities, the study of the developed prototype in everyday life conditions, on exploring how blind people learn to use new strategies to improve their perception of the environment and the exploration of the possible cortical brain areas involved in this process, using functional imaging techniques.

1. Introduction

This work describes the results obtained in a multidisciplinary project called Virtual Acoustic Space (VAS). We have developed an electronic prototype oriented to sensory substitution and this device makes it possible to study the human auditory capability of perceiving complex spatial patterns using a three-dimensional matrix of virtual sound sources.

The basic idea of this prototype can be intuitively imagined as trying to emulate, using virtual reality techniques, the continuous stream of information flowing to the brain through the eyes, coming from the objects which define the surrounding space, and which is carried by the light which illuminates the environment. In this scheme two slightly different images of the environment are formed on the retina, with the light reflected by surrounding objects and processed by the brain to generate its perception. The proposed analogy consists of simulating the sounds that all objects in the surrounding space would generate, as these sounds are capable of carrying enough information, especially the source position, which would allow the brain to create a three-dimensional perception of the objects in the environment and their spatial arrangement, after modelling their position, orientation and relative depth.

This simulation generates a situation which is equivalent to covering all surrounding objects (doors, chairs, windows, walls, etc.) inside a frontal perception field with small loudspeakers emitting sounds depending on their physical characteristics (colour, texture, light level, etc.). In this case, our objective was to study whether the brain can access this information together with the sound source position, using its natural capabilities. The global perception of all these sounds enable the blind person to perceive and form an idea of what his surroundings are like and how they are organized, up to the point of being capable of understanding and moving in it as though he could see them.

Different devices that offer different levels of environmental information for a blind person's orientation and mobility using sounds have been developed [4, 6, 8, 9]. Some of them pursue an object's shape and spatial pattern recognition [3]. There are also very important works in this field based only on touch (Bach-y-Rita, P., et al., 1969). What we have tried to explore is not so much the possible capability of recognizing spatial features of the objects using sounds, but whether a sensation of perceiving a whole picture of the object, present at a time and extending in 3D space can really be generated, as would be described in the subjective experience of a visual image [14, 5]. Our system aims to provide its users with an orientation capability in the nearby space by building a perception of the surrounding space itself at a neuronal level [5, 14], which could be used by the blind person not only as a guide for moving, but also as a way of creating a brain map of how their environment is organized.

Bearing in mind that aim, we first explored the possibility that the blind person could perceive the spatial features, basically shape, width and location, of single objects like flat or linear shaped surfaces. The surrounding space could be simplified in these acoustic objects allowing the blind a better orientation and mobility.

2. Material and Methods

The portable system includes two micro-cameras on each side of a pair of spectacles, see figure 2. The images captured by these micro-cameras are processed using different possible types of artificial vision algorithms (geometric feature detection, stereo-vision, etc) [2]. The 3D-matrix calculated by the above mentioned device is called the *depth map*, which is the information about the position in 3D space of every x,y,z boxel coordinate of the surfaces of the objects present in the scene (figure 1, part b). The depth map generated by the visual system is sent to the acoustic subsystem, where it reproduces a random sequence of short sounds, like “clicks”, one for each of the positions previously registered in the depth map. Every sound has been previously auralized (processed in such a way that when perceived through headphones it seems to come from a certain position in the environment).

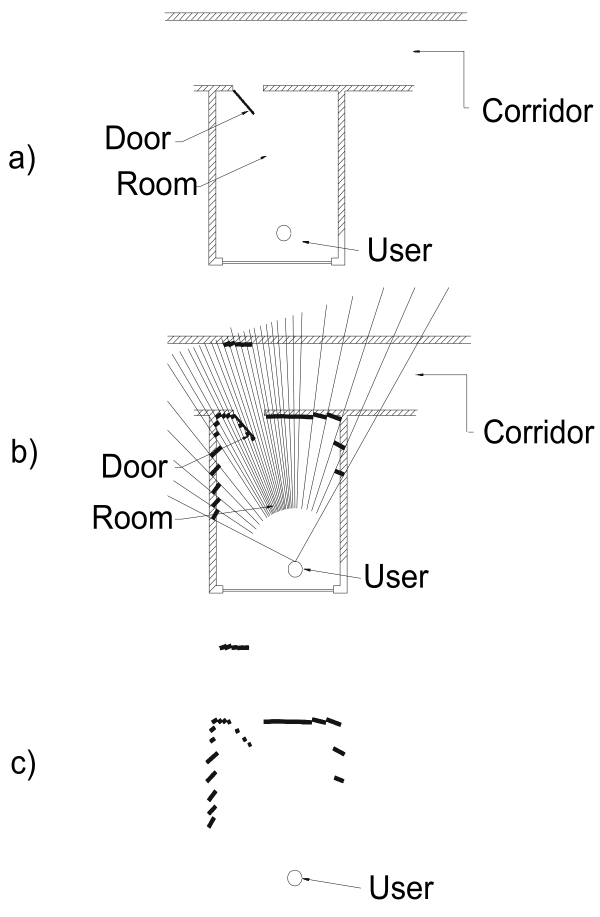

Figure 1.- Two-dimensional conceptual diagram describing the virtual acoustic space construction.

The sound map is then calculated from every captured boxel from the object surface according to height, azimuth or horizontal position, and distance (figure 1, part c). In the present prototype a short and impulsive sound, without pitch has been used to encode every position. This type of sound is easily locatable. The musical effect of hearing this stimulus could be described as perceiving a large number of rain drops striking the surface of a pane of glass. A field of 80 º in the horizontal axis by 45 º in the vertical one is spatially resolved into 17 (horizontal) by 9 (vertical) by 8 (distances from subject) coordinates or pixels, which, because of the third dimension of depth, we usually refer to them as “s*tereopixels*”. In order to explore the practical potential of this kind of stimulus for the blind person, preliminary studies have just begun with a the latest portable and just out of the laboratory functioning version of the VAS, with a resolution of 31x23 steropixels and 16 distances from the subject and the possibility to encode up to 8 grey levels.

When a person is in front of a particular scene, they receive an acoustic input consisting of a set of auralized “clicks", with a randomized order of emission, corresponding to the calculated 3D coordinates of the objects. This set of “clicks” is sent to the person in a time period of 153ms, after which the next acoustic image is sent. Depending on the number of coordinates that the objects occupy inside the perception field, there is a variable interclick interval, never less than 1 ms.

To make it seem that a sound sent through headphones comes from a certain point in space (auralization), the sound is processed off line with the so called Head Related Transfer Function (HRTF) technique [15]. This is a function which includes the effect that the head and the rest of the body have on any sound wave that comes towards them. This effect is different depending on every direction for each ear. These functions are calculated for every user. They are obtained from the signals recorded by a pair of small microphones located at the entrance of a totally-blocked ear channel. A sound wave coming from a loudspeaker located in every position to be simulated, must be recorded.

For each position in space, a set of two HRTFs is needed, one for each ear, so that the interaural time and intensity difference cues, together with the behaviour of the outer ear are taken into account. In our case, we are using a controlled-reverberating environment, so the measured impulse responses also include information about echoes in the room.

A two-dimensional example of the way in which the prototype can work to perform the desired transformation between space and sound is shown in Figure 1. A current limitation of the artificial vision is that the surfaces must have enough texture or presence of linear components to allow the stereovision algorithm to obtain the depth map. In the upper part there is a very simple example environment, a room with a half open door and a corridor. The user is standing near the window, looking at the door. Drawing b, shows the result of dividing the field of view into 32 *stereopixels* which actually represents the horizontal resolution of the vision system, providing more detail in the centre of the field in the same way as human vision does. The description of the surroundings is obtained by calculating the average depth (or distance) of each *stereopixel*. This description will then be virtually converted into sound sources, located at every *stereopixel* distance, thus producing the perception depicted in drawing c, where the major components of the surrounding space can easily be recognized (the room itself, the half open door, the corridor, etc.)

This example contains the equivalent of just one acoustic image, constrained to two dimensions for ease of representation. The real prototype will produce about ten such images per second, and include a third (vertical) dimension, in our hypothesis this frame rate is fast enough to construct a cerebral visual-like perception of the surroundings.

In summary, two completely different procedures are needed for the implementation of a system capable of performing this simulation. First, it is necessary to capture information of the surroundings, basically a depth map with simple features such as colour or texture. Secondly, every depth has to be converted into a virtual sound source, with sound parameters coherently related to the features and located in the spatial position contained in the depth map. All this processing has to be performed faster than the speed of human perception, i.e. approximately ten times per second.

Methodologically, our main interest was to elucidate whether an auditory 3-D image can be perceived by using the designed acoustic stimuli. Preliminary studies in the first stages of the project showed that we can generate an externalization effect and an accurate localization performance of virtual sound sources emitting the selected sound stimuli (Dirac delta) and blind people detected the position of the source with more accuracy than people with normal vision (controls). These results have been observed in azimuth, elevation and distances from experimental subjects [5], in the same way as has been referred to by other authors in the corresponding literature, [15]. The individualized HRTFs for 1003 positions in the frontal field were measured in 5 blind people and several training sessions were held, where, different stimuli and perceptual tasks of increasing difficulty were carried out, such as externalization, distance perception, localization of a punctual sounding area, and perception of non punctual sounding surfaces, like walls and lines of different widths. In these sessions, the blind volunteers learned to focus their attention on exploring the presence or absence of an auditory spatial image with non common spatial qualities such as a possible impression of extension through the perceptual field, with external limits defining a rudimentary shape.

To obtain a non subjective result about the explored ability a sequence of different figures was presented to each subject consisting of 50 cm lines oriented in the horizontal, the vertical or the diagonal axis, versus, punctual acoustic areas located at the ends of the vertical and the horizontal lines and in the centre of them. They were presented at a distance of 90 cm. The main objective was to explore whether the blind person could perceive the presence of a linear bi-dimensional component instead of a punctual one, in the auditory spatial image experience.

The subjects were asked to pay attention to the perceptual aspect “where the sounds seem to come from”. The verbal responses describing the spatial image perceived and the coordinates signalised or drawn in the air by each person were collected using a magnetic 3D-tracking device. At the end of the experimental session, a sighted person moved the extended arm of the blind subject to signal the correctly shown figures, emulating the movement that the blind person would make if they could actually see them.

3. Results and discussion

All the afore mentioned experiments have been carried out with blind people of different ages, blindness onset, duration and degree of severity. Most of them have good hearing and orientation in familiar environments.

The spatial features of these objects, when presented to the subjects, can be correctly described by a large majority of them. Only a few people had difficulty in perceiving the three-dimensional correctly externalized sounds. Regarding the quality of their subjective experience, they said they could perceive all the presented object, perceiving a kind of figure against the background. In the first acoustic object presentation, this happens after a short exploration period of about 1 or 2 minutes. In the following presentations, the figure perception is more immediate. It is possible to distinguish the presence or absences of discontinuities in the spatial extension of the figure. The capacity to recognize the surrounding space was explored using a dummy room, with 4 walls including a table, a window and a column, easily modifiable to prevent the memorization of the object´s location. Figure 3, part A, shows the schematic representation of the above-mentioned room, with a particular object distribution. Part B, is a drawing by a blind person using the developed prototype without relying on touch.

For every individual figure presented the auditory signalized coordinates were compared with the control one (visually obtained). The criterion used to decide whether a line was perceived as opposed to a punctual figure was that at least 3/4 of the figure size was signaled and the verbal response was coherent with the real figure. Concordance in shape (linear versus punctual), orientation and width was evaluated by using information from the graphs containing both drawings. As a representative example in figure 3, parts C and D show

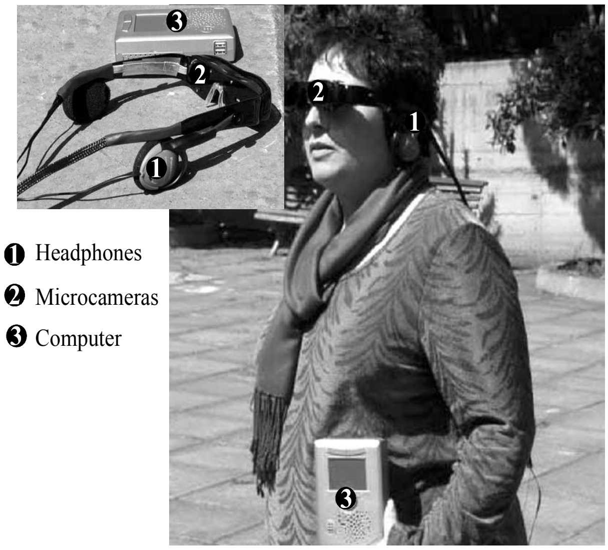

Figure 2.- Photography of the portable prototype. The numbers shows the principal components.

some examples of the drawings comparing the one drawn by the subject using the virtual acoustic device (dotted) and a visually drawn reference one (line).

The orientation of the figure was not mistaken by blind people, except for a few cases when some people signaled a horizontal line when a diagonal one was presented.

In figure 3, part E shows the results obtained from 5 blind people in response to the spatial extension of different acoustic linear figures (similar to C and D, in the above mentioned graphs) as a percentage of accuracy. The obtained results were considered acceptable, taking into account the low resolution of the used device. Part F of the same figure shows the perception error measured in cm, using the tracking device, from the 5 blind subjects mentioned above, in response to the figure dimension. As can be seen in this figure, a previous short period of learning (only few hours) seems to be very important to reduce the committed error, but further studies with a wider sample are necessary. Therefore, we have observed that the perception of a sounding area extended within certain limits in the space takes place almost immediately, as opposed to other stimuli used in sensory substitution [1, 6] which require a long period of learning.

The results suggest that it is possible to perceive the presence, the position and the dimensions of the previously non sounding detected objects from an acoustic stimulus, at least when single objects are presented.

It is not that it is possible to locate many sound sources in different places, and then deduce what the shape and other spatial features of a supposed object might be. We believe this is perceived as a whole object, extending in the surrounding space with its defined spatial features.

This perception seems to be accompanied by an impression of reality, like a vivid sensation of the presence of the object we have attempted to reproduce. It

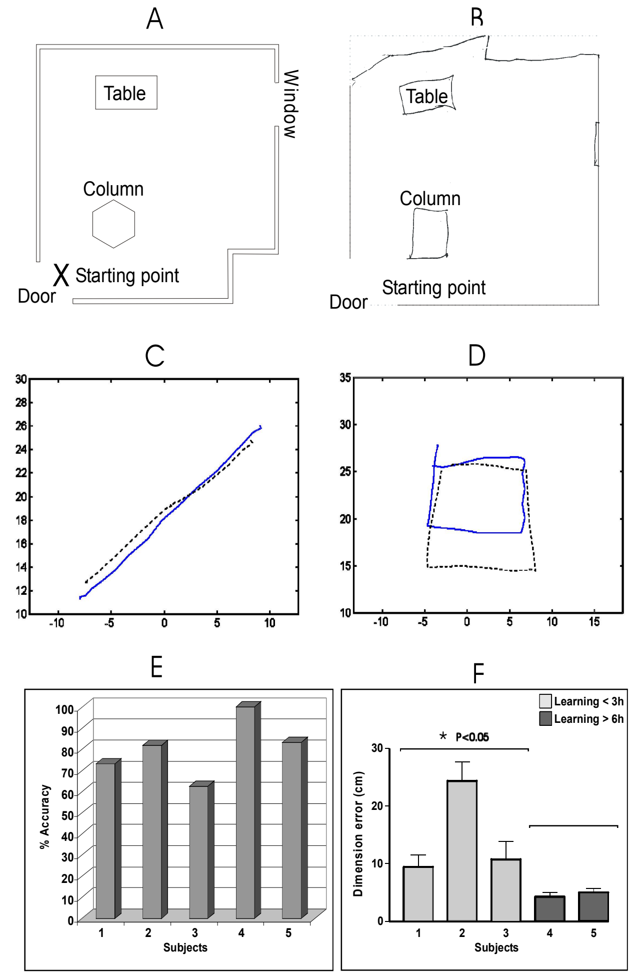

Figure 3.- Part A, B, some figures drawn from visual (continuous blue line) and auditory (dotted black line) information recorded from a person wearing the VAS device when signaling at a continuous line and a rectangle. Part C, D, schematic representation of the experimental room, with a particular objects distribution. B. Drawing made by a blind person after a very short exploration, using the developed prototype, without relying on touch

might be interesting to mention that, in some subjects, the three-dimensional pattern of sound-evoked perceptions had mental representations which were subjectively described as being more similar to the visual images than to the auditive ones.

It is important to mention that in several blind people the sound perception of the environment was accompanied by a simultaneous visual evocation. Therefore, some of the blind people, and even one sighted person during a migraine episode, said they perceived, luminous sparkles (described as small lights or stars) coinciding with the auditory sensation in exactly the same spatial location where they located each sound source. There are a few precedents of this strange phenomenon in the literature [10, 16], known as phosphenes induced or evoked by sounds, but here it appears in their correct spatial location, and interestingly this phenomenon lasts for many years, at least until the present. These punctuated spots of light (phosphenes), are always located in the same positions as the virtual sound sources. Phosphenes did not flicker, so this perception gives a strong impression of reality and is described, by the blind, as visual images of the environment. This seems to depend on both individual factors and on the special nature of the stimulus, and we are attempting to elucidate this by using functional magnetic resonance imaging (fMRI).

Another interesting question that we are studying is related to the ability of most blind people of developing their other remaining senses (hearing, touch, etc.) more than those of sighted people. This is and has been a very important question of debate for a long time. Anecdotal evidence in favour of this hypothesis abounds, and a number of systematic studies have provided experimental evidence for compensatory plasticity in blind people, [12, 13, 17]. Other authors have often argued that blind individuals should also have perceptual and learning disabilities in their other senses; such as, the auditory system, because vision is needed to instruct them [18]. Thus, the question of whether intermodal plasticity exists has remained one of the most vexing problems in cognitive neuroscience. In the last few years, results of PET and fMRI in blind people indicate activation of areas that are normally visual during auditory stimulation [19] or Braille reading [17]. In most of the cases, a compensatory expansion of auditory areas at the expense of visual areas was observed[12]. Similar results have been observed by our group.

We have explored, using fMRI [19], cerebral areas involved in auditory single sound, virtually externalized to one metre away and 45 degrees in the horizontal plane to the right. So far, 6 sighted female subjects and 4 blind females (2 congenital and 2 late onset) have been scanned while presenting a rest-stimulus paradigm consisting of a repetitive train of clicks perceived in a single point, inside the head (monoaural) and outside the head (externalized virtual position). In the blind subgroup a relative shift of the areas activated with the sound stimulus towards more posterior areas (temporal-occipital) which are involved in sighted people on visual processing has been found. Concretely, besides Heschl-planum temporal bilateral activation with apparent left predominance, activation at the bilateral temporal-occipital union, areas 37-19 also appears.

Our results support previous findings, particularly in congenital blindness [5], of an involvement of occipital areas in cross-modal auditory processing, and suggest a similar response in late onset blindness can be found following on from the hypothesis that the visual cortex remains functional in peripheral blindness, which could be understood as being caused by a plastic reorganization, as well as, by functional recruitment. Current studies are directed at continuing the verification of the parietooccipital area´s involvement in perceiving the surrounding acoustic stimuli, both in early and in late onset blindness.

Accepting that stated above, these findings pose several interesting questions: What is the kind of perception that a blind individual experiences when a “visual” area becomes activated by an auditory stimulus? Does the co-activation of 'visual' regions add anything to the quality of this sound that is not perceived normally, or does the expansion of auditory territory simply enhance the accuracy of perception for auditory stimuli?.

To answer these questions neurobiological research, including more studies of functional neuroimaging, on the above-mentioned subjects, needs to be carried out.

Enhancing the non visual abilities of blind people to levels similar to those of sighted people is a highly complex and difficult task because of the vastly superior capacity of the visual channel to process information. Nevertheless, this prototype can provide partial compensation for the lost function by increasing the spatial information coming through the auditory system.

Our objectives are now focused on: a better delimitation of the observed capabilities, the study of the developed prototype in everyday life conditions, on exploring how blind people learn to use new strategies to improve their perception of the environment and the exploration of the possible cortical brain areas involved in this process, using functional imaging techniques.

Acknowledgements

This work was supported by Grants European Commission – infso; FP6-027063 and : FIT-350300-2005-092 from the Spanish Ministry of Industria, Turismo y Comercio

4. References

[1] Bach-y-Rita, P., Collins, C. C., Saunders, F. A., White, B. and Scadden, L. (1969) “Visual substitution by tactile image projection”. *Nature*, 1221, 963-964.

[2] Begault, D. R. (1989), “3-D sound for virtual reality and multimedia”. 1994. AP Professional. Benjamin, J. M. (1974) “The laser cane”. Bulletin of prosthetics research, veterans administration, Washington

[3] Capelle C., Trullemans C., Arno P., and Veraart C. (1998) “A real time experimental prototype for enhancement of vision rehabilitation using auditory substitution”. *IEEE Transactions*

on Biomedical Engineering. Vol. 45, nº 10, pp 1279-1293.Oct.

[4] Dodds, A. G., Armstrong, J. D. and Shingledecker, C. A., (1981) “The Nottingham obstacle detector: development and evaluation”. *Visual Impairment and Blindness*, 75, 203-209.

[5] González Mora, J.L., Rodríguez-Hernández, A., Rodríguez-Ramos, L.F., Díaz-Saco, L., Sosa, N. (1999). “Development of a new space perception system for blind people, based on the creation of a virtual acoustic space”. Proceedings of the International Work Conference on Artificial and Natural networks. Vol. 2, pp 321-330. Springer.

[6] Kay, L. “Electronics aids for blind persons: an interdisciplinary subject”. IEE review. IEE Proceedings, Vol. 131, Pt. A, No. 7, 559-576. September 1984. Lessell, S. and

[7] Cohen M.M., (1979). “Phosphenes induced by sound” *Neurology*, V 29, 1524-1527. Locke, J. (1991). An Essay Concerning Human Understanding (Reprinted 1991, Tuttle).

[8] Loomis JM, Golledge RG, Klatzky RL, (1998). Navigating System for the Blind: Auditory Display Modes and Guidance,. Presence, 7(2), 193-203.

[9] Loomis JM, Klatzky RL, Golledge RG, (2001). Navigating without vision: basic and applied research. Optom Vis Sci. May;78(5):282-9. (Loomis JM, et al., 2001)

[10] Page, N. G., Bolger, J. P. and Sanders, M.D. (1982), “Auditory evoked phosphenes in optic nerve disease”. *Journal of Neurology, Neurosurgery and Psychiatry*. V. 45, 7-12.

[11] Rauschecker JP, Korte M. (l993.) Auditory compensation of early Blindness in cat cerebral cortex. *Journal of Neuroscience*, 13(10) 4538:4548.

[12] Rauschecker JP. (1995). Compensatory plasticity and sensory substitution in the cerebral cortex. *TINS.* 18,1,36-43.

[13] Rice CE (1995) Early blindness, early experience, and perceptual enhancement. *Res Bull Am Found Blind* 22:1-22. Rock, I., (1966). The Nature of Perceptual Adaptation. Basic Books.

[14] Rodríguez-Ramos L.F., Chulani H.M., Díaz-Saco L., Sosa N., Rodríguez-Hernández A., González Mora J.L. (1997) “Image And Sound Processing For The Creation Of A Virtual Acoustic Space For The Blind People”. *Signal Processing and Communications*, 472-475.

[15] Rodríguez Hernández, A; Sosa, N , Rodríguez Ramos, L.F. , Chulani, H. Díaz Saco, L, y González-Mora, J.L.. "Percepción del entorno en personas ciegas a través de un estímulo sonoro espacial virtual generado por computador". Actas del Congreso Iberoamericano 3º de Comunicación Aumentativa y Alternativa, 1º de Tecnologías de Apoyo para la Discapacidad, (Oct. 2000). pp 81-84. Ramón Cerés y Comité Organizador Editores.

[16] Rodríguez Hernández, A.F., González Mora, J.L, Pujol, J., López García, J.A., Muñiz Montes, J.R., Rodríguez Ramos L.F., Chulani, H.M. and Burunat, E., (2003). “Functional Magnetic Resonance Imaging of spatial auditory perception in blind and sighted human subjects” Accepted to be presented at the XXXII Spanish Society of Physiological Sciences Congress.

[17] Sadato N. Pascula-Leone, A. Grafman, J., Ibáñez, V., Daiber, M.P., Dold, G., Hallett, M. (1996). Activation of primary visual cortex by Braille reading in blind people. Nature. 380,526-527.

[18] Veraart C., De Volder, A.G., Vanet-Defalque, M.mC., Bol, A., Michel, Ch., Goffinet, A.M. (1990) Glucose utilisation in visual cortex is abnormally elevated in blindness of early onset but decreased in blindness of late onset. Brain Res. 510, 115-121.

[19] Wightman, F. and Kistler, D.J. (1989)“Headphone simulation of free field listening I: stimulus synthesis”. Journal Acoustical Society of America. V. 85 858-867.