The Role of Astronomy in Society and Culture
Proceedings IAU Symposium No. 260, 2009
D. Valls-Gabaud & A. Boksenberg, eds.

doi:10.1017/S1743921311002845

Virtual Acoustic Space: Space perception for the blind

Luis F. Rodríguez-Ramos

Instituto de Astrofísica de Canarias,
Calle Vía Láctea, s/n, 38200 La Laguna, Tenerife, Spain
email: `lrr@iac.es`

Abstract. This R&D project implements a new way of perception of the three-dimensional surrounding space, based exclusively in sounds and thus especially useful for the blind. The innate capability of locating sounds, the externalization of sounds played with headphones and the machine capture of the 3D environment are the technological pillars used for this purpose. They are analysed and a summary of their main requirements are presented. A number of laboratory facilities and portable prototypes are described, together with their main characteristics.

Keywords. Sound processing, sound externalization, Head Related Transfer Function, machine vision

1. Introduction

After a long history of collaborations in technology, the Institute of Astrophysics of the Canary Islands (IAC) and the University of La Laguna (ULL) started in 1995 the Virtual Acoustic Space project, a highly multidisciplinary effort towards the perception of the surrounding space using sounds. The Technology Division of IAC traditionally devotes some percentage of its man-power to projects not directly related to astrophysics, basically providing support and access to high level technology to nearby located research groups, mostly in medicine and related fields.

The possibility of building a perception of space for humans using only sounds came from the innate capability of people to locate the relative position of sound sources, and had a direct and obvious application for blind people. A number of researchers of very different areas, from psychology to real time hardware engineering, blind and normal sighted, including several physicists and even a sculptor, were put together to develop this project and still continues on work towards the objective of doing research and development in the perception of space using sounds.

A huge number of institutions and funding agencies have joined and/or provided support to this project along the years, from the local Government of the Canary Islands to the Frame Program VI of the European Union, adding up more than 4 M€ of direct funding, excluding overheads and the cost of the researchers with permanent position. Direct contributions of the associations of blind people from Spain, Italy and Germany should also be mentioned. The patent P86002283, held by IAC and ULL, protects the intellectual properties involved in the EAV project with regard to possible commercial uses.

The three main pillars of the project are described below, together with the most relevant results: The capability of spatial identification of the source of a sound, the externalization of a sound and the machine capture of the surroundings of the subject. The combination of all three obtains the illusion of an auditory image of the 3D space that can be 'felt', in the same way that happens with vision.

2. The capability of humans to locate sounds

Many animals, mostly mammals, have very sophisticated sound perception capability. A number of them rely basically on this sense for environment perception, like bats, but this is not the case of human beings. However, locating sound sources is an innate capability of humans, and has been extensively studied in the literature. Human brain builds the location of the source of a sound using a number of clues, unconsciously analysed, providing separated information about the three head-based coordinates and combining them with a priori knowledge to generate the perception of the position of the sound source.

The azimuth clue is the most accurate and it is based on the existence of two separated sensors, two ears located at each side of the head. The time of arrival difference between both ears, and the intensity difference generated by the occlusion of the head is a powerful piece of information difficult to be fooled, allowing for an accurate positioning in the horizontal plane. Things are not so easy in the elevation coordinate, because both ears are located at the same level with respect to the horizontal plane†. But clearly our outer ear is not symmetric with respect to the horizontal plane, and thus the 'processing' made to a sound coming from above the line of sight is very different than from below. Of course there is still the lack of knowledge about the original nature of the sound, but the brain learns the nature of the processing and de-convolves it with reasonable accuracy. This processing is mainly concentrated within the higher part of the audible spectrum, above 8 KHz, where the wavelength of the sound in air is comparable with the physical size of the outer ear.

The most difficult information to be inferred is the distance from the source to the subject. There is an obvious clue regarding the perceived sound level, plus some low-pass filtering depending on the propagation, and also the time of arrival of the direct path with respect to wall echoes. This information is poor in general and unfortunately the sound distance is not located accurately. The nature of the sound plays a very important rôle in the process of source location. Sounds with a high spectral content provide much richer information to the brain, allowing phase delay matching at many frequencies, or de-convolution information big enough to compute the elevation angle of arrival of the sound. It is remarkably true that the sound normally used by people to ask for the attention of other person is something like two very short 's' sounds, basically white noise with two leading edges that allows maximising the chance for evaluating time differences and phase matching. On the contrary, a pure tone reproduced at the laboratory with a loudspeaker is virtually impossible to locate: It really seem to be coming from everywhere.

After the completion of the study of the available bibliography, we decided to carry out three vital experiments with both normal sighted and blind people, oriented to decide the sound to be used for building the auditory image. We built the one-source experiment ('*Monofuente*') where the accuracy of sound location with respect to its nature was analysed, then the two-source experiment ('*Bifuente*') to evaluate the capability of identifying separately two sounds presented almost simultaneously, and finally the multi-source ('*Multifuente*') where a complete shape built using sounds was presented to the subject using a matrix of 8×7 real loudspeakers. With the results of these experiments, we decided to use the sound which maximised the amount of spatial information provided to the brain, a huge number of short *chirps* (pink-filtered Dirac's deltas) with very high spectral content, lasting for a few tens of milliseconds but starting with one millisecond delay.

† This is not the case of other animals, such as the Barn Owl.

https://doi.org/10.1017/S1743921311002845 Published online by Cambridge University Press

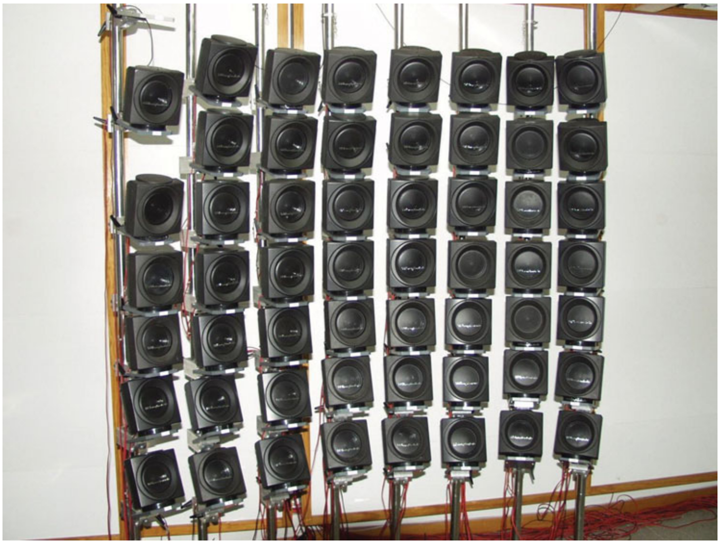

Figure 1. Array of real loudspeakers used during the multi-source experiment. The sound designed to maximise the amount of information gathered to the brain is reproduced in some of the loudspeakers, drawing lines or simple shapes to be perceived by the subject.

3. The externalisation of sounds

The second pillar of the Virtual Acoustic Space project is the externalisation of sounds. This concept was originally developed within virtual reality research and basically means that a sound which is reproduced using headphones is perceived by the subject as if it were coming from somewhere in front of him/her, at a certain distance. The use of this perception 'fooling' is clearly vital for the practical use of the auditory image capability, because it is not viable to cover every object with real sound sources, but it can be done virtually with fairly high degree of success.

In order to fool the brain and virtually generate the perception of sound coming from somewhere in the room, we found it was enough to reproduce the clues used by the brain to locate the sounds, as described previously. These clues can be modelled as a consequence of the sound propagation from the sound source to both inner ears (eardrums), and formulated as two Head Related Transfer Functions (HRTFs), one for each ear, that can be measured and simulated by digital processing on the sounds, before being reproduced by the headphones. HRTFs are specific for every subject and its accurate measurement requires the insertion of a tiny microphone at the ear canal, and to register the processed sound generated by a real loudspeaker and propagated to the inner ear of the subject. We devised a number of approaches to this task, crucial for the success of the project, but affected by a number of practical problems, like the amount of time required for the measurements and the comfort of the user during the measurement. An automated robot with six degrees of freedom was built for simplifying the measurement runs, but in parallel a model of the subject was built using sculptural techniques, being specially accurate in the reproduction of the outer ear and a fairly copy of the person's face. We found that measurements made with mannequins were very similar to the ones made with the real person, and all problems related to the user comfort were solved this way.

However, to overcome many practical limits encountered in the accuracy of the measurement in relation with the generation of the auditory image, we found viable to train

https://doi.org/10.1017/S1743921311002845 Published online by Cambridge University Press

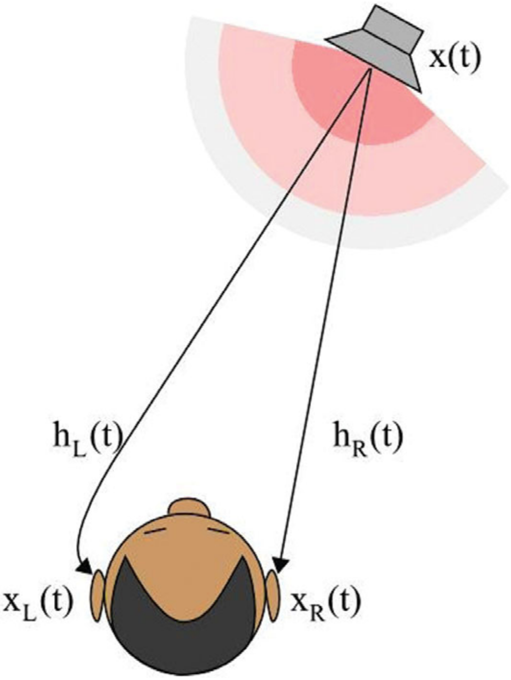

Figure 2. Mathematical formulation of the Head Related Transfer Function, which described the sound propagation process from the source to the eardrum, thus including all clues used by the brain to locate sounds Credit: Inkscape.

the user's brain to use 'standard' HRTFs, instead of their own, in less than two hours time when using an adequate protocol, obtaining fairly realistic externalised sounds.

4. Capture of the shape of the environment

The third pillar to support the generation of the auditory image of the environment is precisely the capture its shape, in order to allow the system to generate the sounds accordingly. A lot of research has been done within this field, and a number of systems have been made available commercially, but there is still a lot of work to be done towards the computer capture of the surroundings, a task made very easily by the human brain but still unsolved to machines in a general outdoor case. A number of techniques have been used to provide information about the shape of the user environment. The problem can be formulated as the measurement of the distance to the closer object situated in front of the user, i.e. a depth map, by analysing an aperture angle of about 100 degrees in azimuth and elevation coordinates. Computer vision allows examining the apparent size of a known item (i.e. a black circle 15 cm diameter) in the image, and directly computing the distance, but this can only be used in laboratory where every object can be covered with instances of the known item. This technique, though simple, is very reliable and easy to implement.

A second and more general approach is to use stereovision techniques to evaluate the distances to the objects. This approach is the one used by human brain to build the model of the surroundings using both eyes, as has been thoughtfully studied by many researchers and developers. It is based in analysing the difference in the position of the object in the image corresponding to each eye, which is directly related to distance by

https://doi.org/10.1017/S1743921311002845 Published online by Cambridge University Press

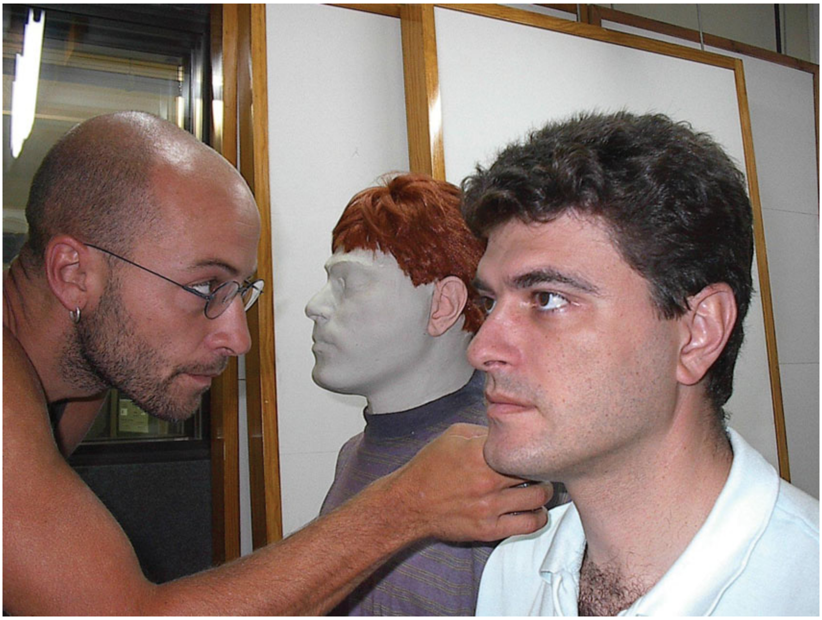

Figure 3. Mannequin built with a very accurate copy of the outer ear, to suppress all practical limits imposed by the user comfort during the HRTF measurement.

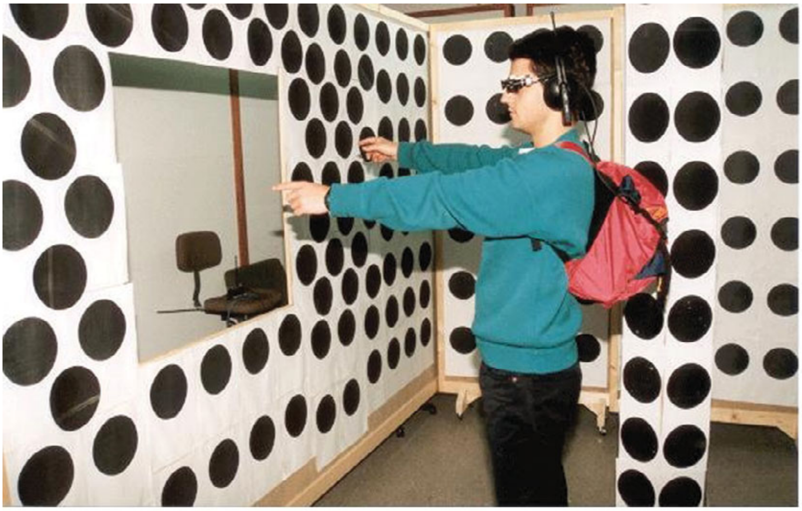

Figure 4. Laboratory environment built using black circles of 15 cm diameter, used to easily compute the distances by measuring the apparent size of the circle images as seen by the camera.

parallax. Unfortunately, the matching problem, i.e. the identification of the same object in both images turned out to be a very difficult one, and occlusions and other practical inconveniencies add high levels of uncertainty and lack of robustness to the technique.

The third technique used is based on the measurement of the time of flight of some beacon back and forth the user and the object. Ultrasounds and laser have been used with success, and the results are reliable if the drawback of 'probing' the environment can be afforded.

https://doi.org/10.1017/S1743921311002845 Published online by Cambridge University Press

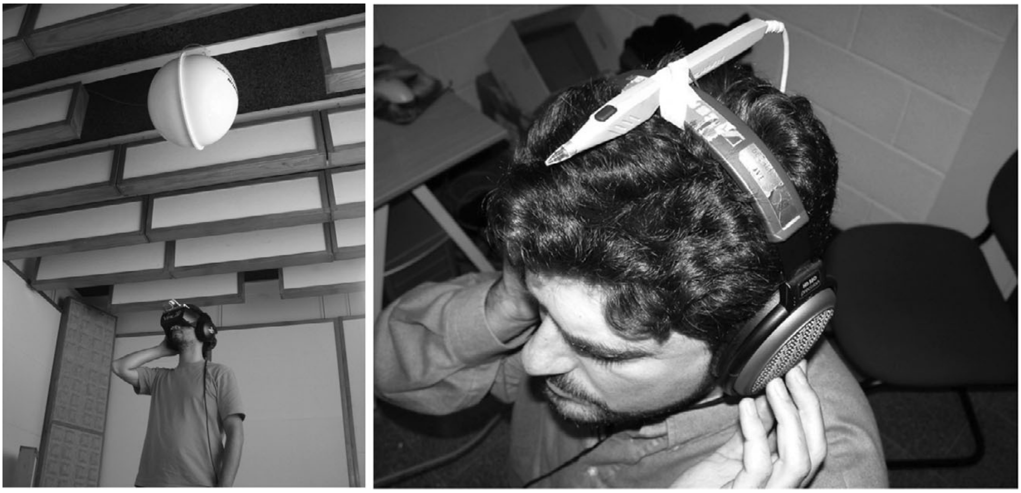

Figure 5. Laboratory system built at ITC, where the user movements are tracked by suitable equipment and a virtual (acoustic) reality is presented accordingly.

Another approach to sort out the problem of surroundings capture is to build a virtual environment within a computer and track the user movements to accordingly present the sounds. This is a laboratory approach very useful to test the accuracy of the virtual world and the response of the user to the stimuli, a research facility of this kind was built at the ITC (Institute of Technology of the Canary Islands).

5. The plenoptic camera

A special mention should be made to the case of the plenoptic camera, a completely new approach to 3D distance measurement developed by a research group at University of La Laguna inspired on the requirements of the Virtual Acoustic Space project, also protected under patents ES200600210 and ES200800126 and ready to be commercialised as a real time 3D camera.

This camera is based on the use of a microlens array located at the focus of a camera lens, providing a huge number of simultaneous views of the scene and then being able of extract distance information, using multi-stereo techniques. This arrangement has also being proposed as a new wavefront sensor for Multi-Conjugated Adaptive Optics for solar images, within the European Large Telescope (ELT).

This development has become a very special case of closing the loop starting on technology generated for astrophysics, then adapting and using it in other fields, and as a result of the further improvements required by the application, serving again for the solution of an important problem in astrophysics.

6. Portable prototypes

A number of working prototypes were developed in order to show the viability of the approach and its application for blind people. The stereovision prototype was presented

https://doi.org/10.1017/S1743921311002845 Published online by Cambridge University Press

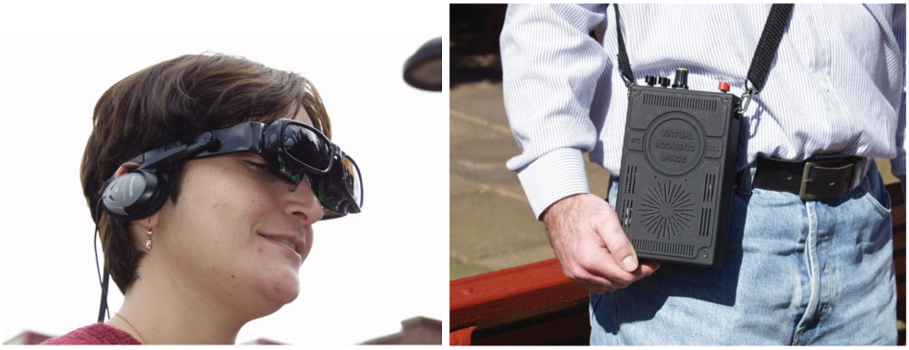

Figure 6. Stereovision based prototype (2002). Based on PC technology, featuring 145 volume elements (17×9×8).

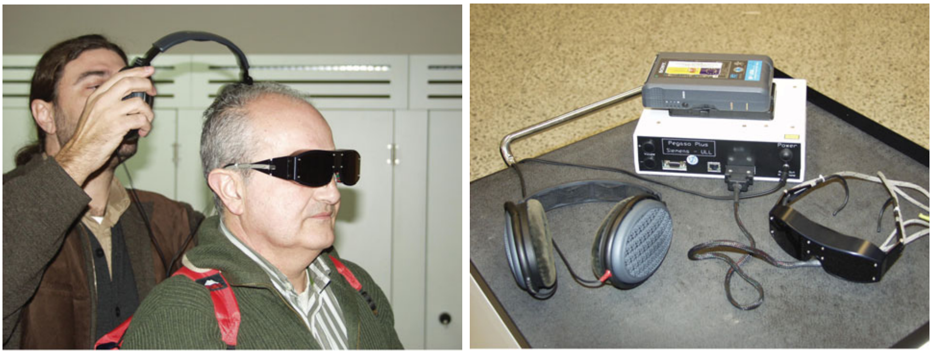

Figure 7. M1 CASBLIP prototype. One-dimensional laser distance sensor provides 64 distances distributed within the horizontal plane at the line of sight level. FPGA based real time adder generates the auditory image.

in 2002 in Los Angeles (USA) at the CSUN meeting. It was based on a PC-compatible portable computer running W2000 operating system. Two webcam-like cameras were mounted in a pair of spectacles and connected via USB to the stereovision software. Sounds were processed off-line and added up in real time depending on the distances measured, and presented to the user by headphones. It featured 17(azimuth) × 9(elevation) × 8(distances) volume elements (voxels).

The M1 prototype was developed by the CASBLIP team, with the funding of the European Frame Program VI. It was based on the one-dimensional time-of-flight sensor developed by SIEMENS, which uses an IR laser to provide 64 highly reliable distances at the level of the line of sight of the user. Sounds were processed off-line with standard HRTFs and combined in real time using FPGA (Field Programmable Gate Array) technology. 10 units were manufactured and distributed to the blind associations of Italy and Germany for evaluation.

Finally, the M2 CASBLIP prototype, will complete the auditory image with information obtained using a number of ways, like stereovision for locating a safe path, or a reading feature capable of identifying any written text in front of the user and reading it to the user through the headphones.

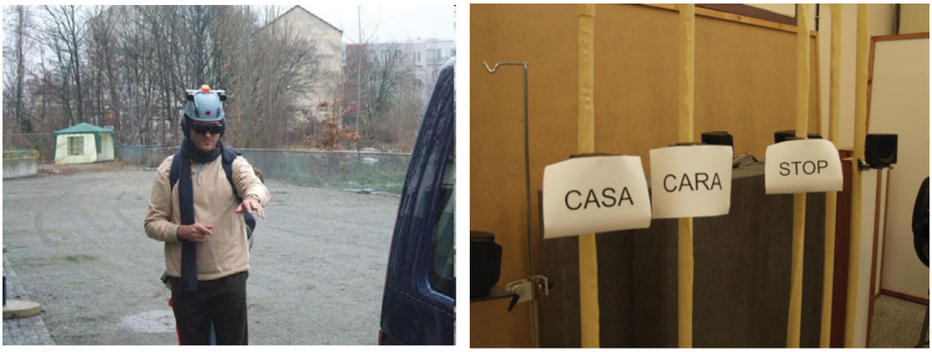

Figure 8. The M2 CASBLIP prototype, which complements the auditory image with information obtained for other sensor and processors, like the reading of texts.

7. Conclusions and future work

After the work of almost fifteen years, it has been clearly demonstrated that shapes can be perceived using sounds, using the human capability of identifying the position of a sound source, provided that proper sounds are used and that the auditory image is built using adequate strategies regarding number, duration and delay between sounds. It can be also concluded that there are a number of procedures to follow in order to obtain the externalization of the sounds when played through headphones, going from the accurate measurement of the HRTF of the user to the training in the use of 'standard' HRTFs. This externalisation can be used for the creation of the virtual acoustic space with success, and has been demonstrated in many users and prototypes.

However, a lot of work is still needed to obtain a reliable environment capture sensor, capable of offering robust measurements outdoors in real life situations. This problem may also be solved by using a combination of complementary sensors, or by improving the smartness of them when identifying and discarding wrong data.

References

Bach-y-Rita, P. *et al.* 1969, *Nature*, 221, 963

Bregman, A. S. 1990, *Auditory Scene Analysis* (Cambridge: The MIT Press)

González Mora, J. L. *et al.* 2003, in *Touch and blindness: Psychology and neuroscience*, S. Ballesteros and M. A. Heller (eds) (Madrid: UNED), p. 371

Kay, L. 1980, in *Animal Sonar Systems*, R.G. Busnel and J.F. Fish (eds) (New York: Plenum Press), p. 769

Rauschecker, J. P., 1995, *Trends Neurosci.*, 18, 36

Rauschecker, J. P. & Korte, M. l993, *Journal of Neuroscience*, 13, 4538

Rodríguez-Ramos, L. F., *et al.* 1997, *Signal Processing and Communications*, 472

Sadato, N. *et al.* 1996, *Nature*, 380, 526

Takahashi, T. T. & Keller, C. H. 1994, *Journal of Neuroscience*, 14, 4780

Wightman, F. L. & Kistler, D. J. 1989, *J. Acoust. Soc. Am.*, 85, 868

https://doi.org/10.1017/S1743921311002845 Published online by Cambridge University Press