

3D Environment Representation through Acoustic Images. Auditory Learning in Multimedia systems.

D. Castro Toledo^{*1}, T. Magal¹, S. Morillas¹, G. Peris-Fajarnés¹

¹ UPV Universidad Politécnica de Valencia, Camino de Vera, s/n. 46022 Valencia, Spain.

Immersive experiences in Virtual Reality environments tend to demand a more realistic Audio-Visual solution. Improving the Audio Perception is a serious challenge to be considered, especially, when an adequate reproduction of sound should significantly enhance the visualization of 3D information.

The translation of the image to sound, under acoustic images, and its inclusion into a Virtual Acoustic Reality promote new cognitive systems, and therefore, a new Language that has to be learnt, trained and practised, in order to understand the meaning of sounds.

This approach opens a broad research line where the sounds can interpret, recognize and localize the objects of a 3D environment, represented by visual information. In this communication, a review of the State of the Art about the new advances in these issues, applications, and the improvements of representation of 3D environments using Virtual Acoustic Reality are presented.

Author keyword: virtual reality; acoustic simulation; cognitive sciences; spatial perception.

1. Introduction.

Developments such as the Internet, multimedia technology, and mobile communication have transformed the way young people spend time outside of school, as Marentette [1] describes in its article. The new disciplinary approaches included in the multi-modal world of Multimedia nearly related and in a continuous feedback are cognitive sciences, education and technology. The sensory perception researchers in auditory and visual techniques offer the knowledge to challenge the multi-modal world questions. In this sense, a narrow communication between each part in order to assimilate the human cognitive perception with the new technologies has to be considered.

Extracted from Marentette [1], “research in the new field of auditory psychoacoustics has led to the development of auditory displays of information that may provide people with additional ways of thinking, learning, and communicating”. Walker [2] points out that “people experience ease of learning and high engagement in the presence of certain sounds and this may impact learning technologies too”, another main aspect of the use of auditory cues in learning activities.

In addition, technology has significantly impacted the lives of young people with disabilities; the representation of images by means of sounds (acoustic images) can help and improve the accessibility to multimedia data by people with disabilities. Besides, the representation of images by sounds enhances both the visualization of Virtual Reality and the immersive experiences.

In this sense, the paper contributes with an overview of the State of the Art in representations of 3D environments by means of acoustic images. A brief summary about these immersive experiences is con-

* Corresponding author: e-mail: macasto@upv.es, Phone: +34 963 879 518

templated in order to localize sound sources and generate the 3D (spatial) sounds, developing virtual acoustic reality. We then specify the features of the auditory perception, contributing in the sensory perception area and offering a helpful tool in the multi-modal world of Multimedia. An explanation of sonification procedures are showed in order to convey the data (or images) to sounds, in our case the image-to-sound mapping. Finally the main aspects of the learning and training of the sounds is described, presenting current discussions in the auditory perceptual field.

2. Immersive experiences in Virtual Acoustic Reality.

The current investigations in virtual reality contemplate the auditory cues slightly. In fact, the visual information is continually enhanced and improved with the new technology available in order to establish new sensations in immersive experiences. But the reality is different, the visual cues only contributes unidirectional information of the real world, opposite to the auditory cues which offer an omnidirectional information of the environment, producing a more natural an inherent sensation of space. In summary, the auditory perception complements the visual perception, according to the cognitive and sensory human perception, for that reason, a conception of virtual acoustic reality has to be taken into account.

The immersion in virtual acoustic reality entails the localization of sound sources and the externalization of these sounds. The sound immersion level scale proposed by Rossi [3], showed in the Table 2-1, describes the methods or techniques that proportionate different immersion sensations. It is based on the sound attributes applied to metrics of immersion capability of spatial audio systems.

Table 2-1 Sound immersion level scale by Rossi [3].

Level	Techniques / methods	Perceptions (results)
0	Monoaural "dry" signal	No immersion
1	Reverberation, echoes	Spaciousness, ambience
2	Panning (between speakers), stereo, 5.1... (n.m. surround multichannel)	Direction Movement
3	Amplitude panning, VBAP	Correct positioning in limited regions
4	HRTF, periphony (Ambisonics, WFS, etc..)	Stable 2D sound fields
5	HRTF, periphony (Ambionics, WFS, etc.)	Stable 3D sound fields, accurate distance and localization

The 3D (spatial) sound allows a listener perceives the position of sound sources, emanating from a static number of stationary loud-speakers or a pair of headphones. In this case, the level 5 reaches the maximum precision in the localization of sound sources, being as accurate as in ideally real world.

In particular, Kapralos [4] describes, in order to localize sound sources, the human auditory system which relies primarily on:

- Interaural Time Difference (ITD): The difference in time between the arrivals of the sound to each of the ears.
- Interaural Level Difference (ILD): The difference in sound pressure level (SPL) between the sounds at both ears.

- Head Related Transfer Functions (HRTFs): The complex interaction of a sound wave with the torso, shoulders, head and particularly the pinna (outer ear) of a listener. Essentially, the pinna of each ear filters every sound wave passing through it in some manner unique to the sound source position. Given these filtered signals, the brain estimates the exact 3D position of a sound source relative to the listener.
- Reverberation: Reflections of the sound waves off of other objects in the environment (e.g. the walls of a room).
- Interaction with Vision: We can determine the location of a sound source which we can see.

3. Auditory perception.

Firstly, a complete description of the Perceptual attributes (dimensions) within audition performed by Hollander [5] is showed:

- Loudness, perceptual sone scale is defined as 40 dB of a 1.000 Hz of tone.
- Pitch, human detects frequencies from 20Hz to above 20.000Hz, although up to 5.000Hz is a true sensation of pitch.
- Timbre, brightness is probably the most well known of the timbral dimensions, is a measure of acoustic energy quantified as the centroid of the perceivable auditory spectrum.
- Frequency is a combination of time and intensity, so auditory perceptual dimensions are all based on time and intensity.

Connecting with the localization of sound sources in immersive experiences, a description of the perceptual considerations of each human auditory system is considered:

3.1 HRTF's functions (monaural cues).

Response azimuth is nearly perfect and response elevation is always a lower in the headphone condition than in free field, by F.L. Wightman and Kistler[6] and [7]. 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. According to Kyriakakis [8], they provide a combined model of the HRTFs for all directions. Achieve reduction in the model size with minimum loss of accuracy.

In the other hand, the acoustic image without Externalization uses azimuth encoded with time delay, elevation with frequency and distance with amplitude (loudness). It is efficient but unnatural to the brain, the solution described by Sodnik [9], comes from divide the visual field into subspaces. (-90° , 90° azimuth / -45° , 90° elevation), 25 points were obtained to position our virtual sources.

Prior measurements, all subjects are introduced into spatial sounds, using this training with three parts according to Sodnik [9] description:

- Random sound playback from azimuth of 0° , 90° , and -90° (intended to introduce HRTFs to the listener).
- Random sound playback from azimuth of 45° , 0° and -45° (intended for the listener to get sense of direction)
- Sequential sound playback from azimuth of 90° , 45° , 0° -45° , -90°
- Sequential sound playback from azimuth of 0° , 15° , 30° , 45° , 60° , 90° .
- Sequential sound playback from azimuth of 0° and elevation of -20° , 0° , 20° , 40° , 60° , 90° .

3.2 Interaural (binaural cues).

According to Kyriakakis [8], low frequencies are better localized by ITD, interaural time differences, and high frequencies by ILD, the interaural intensity differences.

By other hand, Externalization depends on the interaural phases IPD (time adequate relationship) of low-frequency by boundary near 1000 Hz. On interaural level differences in all frequency ranges is equally important, commented by Hartmann [10].

Contrary to the proposal done by Kyriakakis [8], Brungart [11] exposes that Low-frequency interaural level differences are the dominant auditory cue in the proximal region, due to the Distance accuracy which was also found to be dependent on the low-frequency of the stimulus.

The conclusions seems to be equal with the ITD, by low-frequency, but a discussion with the use of high or low-frequency with ILD is presented.

3.3 Distance perception.

The auditory distance cues may potentially play a role in the perception of the distance to a sound source when both the observer and the sound source are stationary, by Kapralos [4]:

1. 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).
5. Type of stimulus used (e.g. familiarity with the sound source).

Loud-ness and reverberation are the two most prominent distance cues. The familiarity of a sound source and its effect on a virtual auditory display is described by Begault [12], where any reasonable implementation of distance cues into a 3D sound system will require an assessment of the cognitive associations for a given sound source.

Commented by Brungart [11], it may be unnecessary to include binaural distance cues (ILD, ITD) in a virtual auditory display. Furthermore, expressed by Blauert [13], the effect of binaural cues on source distance remains an unresolved issue.

3.4. Movement perception.

According to Rossi [3] for the movement perception, the signals should include transients that trigger the precedent effect, e.g. noise burst are better than continuous noise. Two transient and spatially separated sounds occur within short temporal intervals ($< 100\text{ms}$), a single sound image is perceived that continuously traverses through the spatial extent between the two sound sources.

Taking into account the perceptual attributes of the sounds and the elements for the localization of sound sources, according to Hollander [5], it contemplates three types of error in localization:

1. Precision in localization process.
2. Interaural Ambiguity ("cone of confusion" : Front-Back confusion)
3. Lack of externalization of sound due to synthetically spatialized sounds.

4. Sonification.

Sonification is the use of non-speech sounds to convey information. The use of sonification procedures helps the understanding and analysis of data by means of non-speech sounds.

The sonification definition by Kramer [14], “Non-speech sound in computer interfaces can increase the amount of information bandwidth transmitted to the user”, affirms that the non-speech sounds contribute with a major data or information than speech sounds. In fact, the non-speech language is quickly learned each time the users identify a new object. Four techniques involve the main sonification approaches, described by Hermann [15]:

- Auditory Icons: A set of sound pieces, like an auditory sign, which must either be learned or intuitively understood. This method is often used for alarm signals and navigations cues.
- Earcons: Here auditory signs are combined to form more complex messages. Walker[2] defines the earcons like “auditory equivalents of icons which contain a structure that can link them together when information is presented, helping to reduce learning time in multimedia applications”.
- Audification: According to Kramer [14], it is interpreted as a time series which directly controls the audio signal amplitude.
- Parameter Mapping: The employ of a synthesizer where one or more tones are generated choosing sound attributes, e.g. time stamp, duration, volume, pitch, envelope characteristics, brightness, etc...

The last is the most suitable method for the image-to-sound mapping, which corresponding to the translation of the image to sound performed by acoustic images.

A clear example of the use of sonification procedure for educational purpose is described by Upson [16], where subjects receive training in Cartesian graphing over several sessions with sonification software.

5. Discussions.

There are difficulties to match the human perception system in sensitivity or accuracy. The problem now becomes how to read results in the auditory field from human brain. The researcher has to find some way to derive the response. According to Hirvonen [17], there are several listening tests which try to define this response. In general, the listening tests are divided in two main areas: the physical variables caused by test location and implementation and the psychological variables associated with the test subject.

The understanding of sonification procedures involve extended periods of training and learning of the sounds to interpret its meaning. But the problem of the image-to-sound mapping or data -to-sound mapping comes from that all the current methods does not contemplate the principles of psychoacoustic in implementing image to sound conversion methods, according to Matta [18]. In continuous listening tests of the same signal at regular time intervals, although spatialized, produces an unnatural effect and causes a progressive fatigue, experimented by Fusiello [19]. The current technology applied is not sufficient and has to be improved, as Matta [18] comments in his article.

6. Conclusions.

In this article we have presented a revision of the main concepts in 3D sounds, auditory perception and sonification. The inclusions of these new technologies in the auditory field develop virtual acoustic realities have many applications in immersive experiences and sensory substitution.

By itself, the use of sonification procedures in people with disabilities opens new develops in the educational field. In fact, the sonification procedure involves a new language that has to be learnt by the end-users or, in this case, by people with disabilities.

According to the current discussions, the incorporation of new advances in psychoacoustic to resolve the problems of the implementation in sonification and the use of richest spatialized sounds would be able to improve the auditory perception in the future, opening a broad research line.

7. References.

- [1] Marentette, Lynn V., Union County Public Schools (2004)
- [2] Walker, BN & Kramer, G Ecological psychoacoustics and auditory displays: Hearing, groping, and meaning making Online document: <http://sonify.psych.gatech.edu/publications/> . (2004)
- [3] Regis Rossi, A. Faria, Marcelo K. Zuffo, Joao Antonio Zuffo. Improving spatial perception through sound field simulation in VR. VECIMS – IEEE International conference on virtual environments (2005).
- [4] Bill Kapralos, Michael R.M. Jenkin,. Auditory perception and spatial (3d) auditory systems. (2003).
- [5] Hollander, J. An exploration of virtual auditory shape perception. University of Washigton. (1994).
- [6] Frederic L. Wightman and Doris J. Kistler. J Acoust. Headphone simulation of free-field listening i: stimulus synthesis. Soc. Am. 85, (1989).
- [7] Frederic L. Wightman and Doris J. Kistler. Headphone simulation of free-field listening ii: psychophysical validation. J Acoust. Soc. Am. 85, (1989).
- [8] Chris Kyriakakis. Fundamental and technological limitations of immersive audio systems. Proceedings of the IEEE, vol. 86, (1998).
- [9] Jaka Sodnik, Rudolf Susnik and Saso Tomazic. Acoustical signal localization through the use of head related transfer functions. IEEE (2003).
- [10] William M. Hartmann and Andrew Wittenberg. On the externalization of sound images. J Acoust. Soc. Am. 99, (1996).
- [11] Douglas S. Brungart. Auditory localization of nearby sources III. Stimulus effects. Acoustical Society of America, (1999).
- [12] R. Begault. 3D sound for virtual reality and multimedia. Durand National Aeronautics and Space Administration, NASA/TM (2000).
- [13] J. Blauert. Spatial hearing: the psychophysics of human sound localization. MIT Press Cambridge, (1983).
- [14] Kramer, G. Auditory display, sonification, audification, and auditory interfaces. Addison-Wesley, (1994).
- [15] Hermann T., and Ritter H. Listen to your Data: Model-Based Sonification for Data Analysis.
- [16] Upson Robert. Educational sonification exercises: Pathway for mathematics and musical achievement. ICAD02-1 (2002).
- [17] Hirvonen Toni. Headphone listening test methods (2002).
- [18] Suresh Matta, Dinesh Kumar, Xinghuo Yu, Mark Burry. Discriminative Analysis for Image to Sound mapping. ICISIP – IEEE (2004).
- [19] A. Fusiello, A. Panuccio, V. Murino, F. Fontana, D. Rochesso. A multimodal electronic travel aid device. IEEE International Conference on multimodal Interfaces. ICMI (2002).