24th INTERNATIONAL CONGRESS ON SOUND AND VIBRATION
23–27 July 2017, London

DESIGN AND IMPLEMENTATION OF AN AUDIOVISUAL TRACK WITH THE SUPPORT OF AMBIOPHONICS SYSTEM – PART I: PRELIMINARIES

Laura Nataly Villa Arrieta and Natalia del Pilar Farfán León

Universidad de San Buenaventura-Bogotá, Facultad de Ingeniería, Bogotá, Colombia.

Marcelo Herrera Martínez

University of Iceland, School of Engineering and Natural Sciences, Reykjavík, Iceland.
email: marcelo@hi.is

The main goal of the project was to implement 3D sound through an ambiophonics system, with the aim of evaluating its impact on an audio production with advertisement objectives. For that purpose, an advertisement work based on a spatial combination of the narrative elements, voice, music, effects and silence was implemented. The construction of an ambiophonics system is performed for the audio enhancement of the advertisement product. Here, the RACE algorithm is used with the aim to eliminate the acoustic diaphony and to obtain virtual realism. A room is selected for fulfilling with the characteristics of audio playback with the ambiophonics material. There was the need to implement the loudspeaker array for the frontal ambiophonics system, and another for the back one, with the convolution.
Keywords: Ambiophonics, spatial sound, acoustic diaphony, RACE algorithm.

1. Introduction

The search of sound spatiality has been a major concern for many audiophiles and engineers with the aim to create or reconstruct a natural sound environment which may give 3D impression [1], [2].

3D audio may be obtained either with headphones or with an array of loudspeakers. In the former, techniques such as binaural technology, HRTFs, and stereo-panning are used, while in the other, some standards have been developed (Surround Sound, Ambisonics, Ambiophonics and Wave Field Synthesis). The ambiophonics system immerses the listener into the 3D spatial reality, adjusting the individual characteristics of the pinna, minimizing the interaural correlation, generating early reflections and reverberant fields to real impulse responses, eliminating the notch filter due to the diaphony between loudspeakers and using a proper acoustic conditioning of the listening room. In order to facilitate the construction of a new advertisement message, the ambiophonics system is proposed.

Multichannel audio has been a research topic since the middle of last century [3], [4], [5]. In audiovisual films, it appeared in 1940, in Walt Disney´s "Fantasy", which used the ideas of Blumlein – with a special arrangement of microphones- in order to give stereophonic sensation. A breakthrough was observed in the 70´s with the arrival of Dolby Stereo technology, which compressed (analogously as today´s audio compression [6], [7]) in a very convincing way, four channels in space, through a matrix technique. This new system is compatible with monaural cinema sound, being today one of the most popular systems because of its sound and quality in analog cinema rooms. This system is called Dolby Pro Logic in Home Theatre equipment for home, adding a special LFE (Low-Frequency Channel).

One step further was done in 1980 by the Argentinian Hugo Zuccarelli, based on the physical reconstruction of the acoustic field. It is derived from the Huygens principle which mainly states that when the wave is propagating, then the wave front may be assimilated as a secondary source distribution. In recent times a huge number of systems have appeared as Dolby Surround, Ambisonics, Roland RSS and Q-Sound.

In 1992, Dolby Laboratories creates Dolby Digital, originally known as AC-3, or third generation audio coding, a data compression technology which consists of six audio channels, recognized nowadays as 5.1 Surround Sound. In 1993 DTS System (Digital Theater System) was launched by Digital Theater System Inc., consisting of the digital compression and recording of multichannel sound in CD-ROM. After 1995, Sony Dynamic Digital Sound (SDDS) was presented to the world cinema industry as the unique effective format of 8 channels. From 1999 new formats came to the market, adding a central surround channel. These Extended Surround formats have been developed by Dolby as by DTS Technology. In 2000 Dolby Surround Pro Logic format appears in the market, having the possibility to convert any format to 5.1. In 2005, MP3 Surround appears, developed by Fraunhofer IIS (the original developers of MP3) in collaboration with Agere Systems, an advance which makes possible the data transmission and storage with 5.1 multichannel sound.

Due to the fact that the multichannel system is based on the stereo system, it suffers the same problem: the acoustic diaphony. A listener uses both ears, which gives the spatial cues from Duplex Rayleigh theory [1], [8], [9], [10]. The signal crosstalk will give a distortion due to the notch filter formed. In order to seek a solution for this problem, Angelo Farina, Ralph Glasgal, Enrico Armelloni and Anders Torger, in their ambiophonics institute propose a new system called Ambiophonics, with the aim to correct and improve the conventional stereo system and the surround system. Between 2001 and 2010 AES conferences, these authors have presented diverse studies and scientific articles comparing different techniques and psychophysical responses to these stimuli.

In Colombia, the research in this field has been carried on, in 5.1, 7.1 and stereo systems. Lately, the Jecklin Disk recording technique and the ORFT also have been investigated.

A solution to distortion problems is the ambiophonics system, which enables to reduce it, allowing to obtain a solid 3D, clear image with horizontal localization. The achieved sound is clearer and the scenario is wider, at least to 150 degrees in comparison to the 60 degrees' stereo scene. This innovation in the advertisement field can become an important strategy for marketing which may bring to the consumer a different experience while receiving a definite information related to goods consumption.

Initially, the necessary parameters of the room for the implementation of the ambiophonics system are to be gathered. The key part for the implementation in this investigation is the correct performance of the RACE algorithm. For the reverberation time measurement (reverberation time before and after the acoustic conditioning), a dodecahedral omnidirectional sound source was used 01dB OMNI-12, a measurement microphone ECM 8000 from Behringer, an audio interface M-Audio Firewire 410 and the Adobe Audition Software 3.0 for the capture and processing with the Aurora 1.2 plugin.

The implementation of the system was done in the 5.1 audio studio of the San Buenaventura University, with for loudspeakers Adam A8x, Audio Interface M-Audio Firewire 410 and the Pro Tools HD Software for the advertisement audio reproduction with and without the RACE algorithm processing.

A subjective psychoacoustic test was performed to the student population of the San Buenaventura University, evaluating the advertisement impact and the auditory perception in order to precise the differences between the stereo and ambiophonics system.

2. Sound source localization

In order to understand how a stereo system distorts source localization and how the Ambiophonics system corrects it, it is necessary to understand how the brains and the human hearing system localizes sound. Modifications in the signals arriving at our hearing systems due to direction and distance of

the sound event are treated by psychophysics. Auditory pinnae, head, neck shoulders and torso act in a combined manner in this modification. Any alteration of the acoustic signal characteristics, related to the source position is potentially a cue for localization. Cues are often divided into two groups: temporal and spectral, resumed in what is known as Binaural Cues (ILD, ITD) [11].

More recent research indicates that sound source localization synthesis methods based on HRTF (Head-Related-Transfer-Functions) are very reliable while using with headphones [12], [13].

Measurements have widely reconfirmed that human beings are sensible to sound direction [3], [14]. Ambiophonics tries as well to minimize the Acoustic Diaphony, which is explained in the next section.

2.1 Acoustic diaphony

Acoustic diaphony is an acoustic phenomenon which is presented in stereo systems and more complex audio systems as 5.1 and 7.1, among others; it is produced because of listening simultaneously to two loudspeakers with both ears, in other words, due to the fact that a listener does not listen to one single loudspeaker with each ear, but to the couple of them, with two ears.

5.1 and 7.1 systems have even greater acoustic diaphony than the conventional stereo systems, due to the fact, that they use three frontal loudspeakers: left, right and centre. In this case, the left ear will listen to three loudspeakers –as well the right ear- which will make six different versions of the same instrument.

According to Timothy Bock and Don Keele Jr, the interaural difference affects the stereo reproduction mainly in two aspects: its effect in sound images and its effect in frequency response.

Unfortunately, the actual acoustic diaphony in audio systems also generates comb-filter effects, responsible for stereo sound distortion.

2.2 Comb Filter

The comb filter is the main cause of sound coloration or modification, which may be perceived when the sound signal has added reflections or when it is emitted from two or more loudspeakers, located at different distances from our position. This acoustic summation of two similar signals slightly separated in time produces a combined signal which exhibits evident cancellation effects and reinforcement in regular frequency bands.

Considering a conventional stereo system reproducing a soloist localized at the front, both loudspeakers produce similar signals. Left ear listens to the left loudspeaker and about 0.22ms after, it listens to the right speaker, and the comb filter is produced.

With 5.1 or 7.1 systems, the same problem occurs [15].

3. What is a realistic sound reproduction?

Realism is understood as the generation of Hi-Fi sound field in live-sound concerts, games, and films, well enough to satisfy location necessity and auditory perception.

Due to the fact that people are familiarized with the sound perception of a live-sound event when comparing this perception to the stereo system, then the loss of integrity and consistency in a psychoacoustic sound field is perceived.

Human beings have two auditory channels which enable us to achieve realism in playback. Therefore, providing the same acoustic pressure to the entrance of a listener auditory channel is enough, even in the presence of head movements, which he would have been experiencing while the recording session. Everything which is to reproduce –either direct sound or environmental- should come the closest from the correct direction in order to achieve the auditory pathway through a path that passes the normal structures of the pinnae and head. Therefore, only two recorded channels are enough for realistic frontal musical reproduction, in such a way that while reproducing the recording, the cross information is eliminated.

3.1 The Ambiophonics system

Glasgal resumes his method in the following manner: “The ambiophonics system moves the listener to the same space as the musicians, because it accommodates to the individual characteristics of the auditory pinna, minimizing the interaural correlation in the auditory positions, abandoning the traditional equilateral stereo triangle, generating early reflections and reverberant fields of stored impulse responses. It eliminates the crossed information of the frontal loudspeakers in order to simulate concert halls or diverse spaces in which reproduction takes place” [15].

The system uses DSP and a loudspeaker array directly in the front of the listener in order to improve the stereo reproduction and 5.1 surround systems for music, video games or studio applications.

The Ambiophonics system eliminates the acoustic diaphony, uses convolution to create real environments, giving real auditory images in the two-channel playback as LPs, CDs and 5.1 Dolby digital, among others. This, due to the fact that the two-channel recordings do not count with the acoustic diaphony that is produced by the loudspeakers in the usual stereo triangle.

Ambiophonics use ambiopole, a pair of frontal loudspeakers diaphony-free, surround loudspeakers given by impulse response convolutions, room-corrections so to not distort the sound field and the ambiophone - a microphone array conceived for the recording and reproduction – in the most optimal manner.

More advanced versions take into account 5.1, 6.0, 7.1 and Dolby/THX systems in their arrangements as Panorambiophonic system. This one uses four channels as SACD or DVD-A in order to provide a 360º scenario for films or concerts [16].

3.2 Basic ambiophonics system: an ambiopole

In stereo reproduction, the frontal scenario is generated between the loudspeakers; in ambiophonics, the frontal scenario is created from the loudspeakers outwards. An ambiopole consists of a pair of loudspeakers, which form an angle of 20-30 degrees with respect to the listener with a diaphony cancellation system (Crosstalk canceller) implemented through DSP[1].

The physical barrier may be substituted with digital processing. The ambiophonics system has XTC Recursive Ambiophonics Crosstalk Eliminator (RACE). The loudspeakers acoustic boxes are very proximal between them. When getting them even closer, the comb filter is produced at even higher frequencies, and the focusing perception of the virtual central image improves. The more the angle between the loudspeaker acoustic boxes, the worse the localization. This proximity between boxes does not mean a loss in scene amplitude; on the contrary, the physical barrier itself makes that the loudspeakers are un-locatable. Therefore, all the localization is virtual and it is constructed from the binaural signals (ITD, ILD, panning…) included in the recording. The listener can sit in any place between the loudspeaker line, he/she can tilt his head down, etc. but in general the ambiopole only has the capacity for one or two listeners. The surround loudspeakers are fed by ambiance signals which are calculated through convolution, which in terms of an Ambiophonics system is called Ambiovolver. There are different stored impulse responses, from different rooms, churches, and auditoriums around the world, which may be selected according to the type of the recording.

The ambiovolver generates the proper reflections according to the loudspeakers to use and it feeds them with the aim to simulate the acoustic properties of a room or concert hall.

3.3 Ambiophonics Recording

Even though the ambiophonics method can be adapted to the actual stereo material and therefore to the conventional stereophonic techniques, it is better to work with recordings made with ORTF technique or Time Difference Stereo (A/B Stereo Technique). In order to obtain an optimal audio material in the ambiophonics method, the binaural recording must be used, however, there is no need to build an extra external ear. In the ambiophonics method, an external ear similar to the human one must be made.

[1] Digital Signal Processing

With the aim to obtain more exactitude in the ambiophonics reproduction, the recording must be done through an ambiophone which basically is a sphere simulating a head with two omnidirectional microphones located at the places of the auditory channels and protected from the sound at the back part, the sides or above the head by panels.

The ambiophone must capture the direct sound from the scenario and the reflections generated in the audience. Therefore, frontal, lateral and superior reflections should be rejected at the recording point. This implies that the ambiophone must be located in a proper sit in the scenario, generally between the first and the fifth row, so to capture a wide margin in the horizontal plane. These placements are described in [16]. At the playback instance with an ambiopole, the direct sound, and the environmental sound comes from the correct direction (the front) and the ambiopole generates the necessary level and delay values in order to have a realistic localization and spatiality.

With this method, it is possible to reproduce a 150 degrees' scenario, even sometimes till 180 degrees, but with some faults in the extremes that may be solved with the addition of auxiliary loudspeakers related to the ambiovolver. The scene width is wider for the ambiopole/ambiophone than the one for the stereo system/5.1 surround.

3.4 RACE

As said previously, the stereo dipole does not work properly if there is not active cancellation for the loudspeakers. The (Recursive Ambiophonics Crosstalk Elimination, RACE) is an electronic cancellation method that works for two or more channels. At the moment a plug-in VST may be used and its function is to simulate the physical barrier used in Dr. Keele's studios, with the aim to eliminate the acoustic diaphony between loudspeakers.

4. Conditions of the playback system

4.1 Frontal loudspeakers

In an ambiophonics system, the frontal loudspeakers must be located directly in front of the listener with each speaker directed to the listening area.

For better results, the pair of frontal loudspeakers, ambiopole must be the most directional possible. In theory, the ideal loudspeaker for this purpose has to act as a flashlight with the light beam emitted from a single point at the ear level. The more focalized the ambiopole, the more effective the software.

4.2 Ambient sound by convolution

The room impulse response gives us the necessary information (in time and frequency) with the aim to know the signal modifications due to the room.

The existing problem while using three, four or five microphones in the back part of the room and afterwards to record these signals with the aim to capture the proper ambiance of that place.

However, these microphones capture the direct sound and while reproducing it by back or lateral speakers, diaphony is produced, confusion at the auditory pinna and comb filters.

The use of convolution ambiance eliminates the necessity of DTS or Dolby Digital recordings, at least to what it corresponds to music.

4.3 Playback room

In the ambiophonics system, the processing or correction of the reproduction room is not imperative, except the low-frequency region, where, as in stereo, the eigenmodes of the room may cause a problem that may be solved through software for room response correction.

Even though the room correction or treatment is not mandatory for an ambiophonics experience, Ralph Glasgal recommends having certain guidelines with respect to the acoustic conditions of the room in order to generate an ideal listening environment in the ambiophonics reproduction

With the goal to have a realistic room, similar to the experience of a concert hall, any characteristic that would modify the system behaviour in the back and frontal part should be eliminated through the

ambiopole and the convolution. An RT60 close to 0.3 seconds is recommended. The system is designed to be implemented in small rooms or specifically in a regular room.

Reflections, absorption, background noise, must be taken into account. A proper value for the ambiophonics system should be NC-20, as it is usual in recording professional studios. NC-35 is the minimal value for an ambiophonics experience.

5. Audio in advertisement

Advertisement needs besides text, sound, which has to be synchronized to the message, enabling the generation of reactions and stimuli to the listener. An advertisement announces generally consists of about 20, 30 or even 60 seconds, depending on the advertisement intention.

The creative conception of the advertisement message establishes as the main goal, the necessity of articulate expressive ways which enables more original, newer, more impacting and more believable messages.

Moline considers that the advertisement creativity is able to say in an innovative manner or to convert in new what in reality is redundant.

In this sense, the doctoral thesis from Altesch, M., (Statal University of Pensylvania) considers that the originality of the idea and the adaptation of it to the message are the main components of the advertisement creativity.

5.1 Characteristics of an advertisement wedge in the ambiophonics system

The ambiophonics system enables to give a construction to an advertisement from different creative possibilities. Among these distinct features, the following may be cited: 1. It enables a surround space sound until 360 degrees able to generate sensations of similar listening conditions to a natural environment. 2. The system enables to generate 360 degrees' surround sound. 4. It enables multiple source localizations inside any internal point of the created sound field. 5. It enables the recreation of the sound space because a visual referent does not constraint the sound environments. 6. It enables to modify the message based on the knowledge of the listening point and the message itself.

5.2 Duration of the message

It is not possible to construct an advertisement message in 20-30 seconds, which may be the duration of a common advertisement. However, there should exist a direct dependence between the quantitative and the quality aspect, cause the temporal configuration of the message would not influence its creativity or efficiency.

Advertisement efficiency does not grow directly proportional to the size of the advertisement; the memory trace of a 30 seconds announcement is about 2/3 of the 60 seconds one.

5.3 Advertisement

In the advertisement field, there are various forms of announcing messages. They differ by the recorded and mixed instruments, their duration and the medium of transmission, as well as the target, the type of market and the medium in which the product is presented. However, its range may be enhanced while mixing the sound with the image as in TV, cinema or internet. In this manner, may the case that characteristics of advertisement messages have analogies with the audio-visual field. In publicity, the hearing sense has a key role. Among others, the key factors in a good advertisement are the idea, the locution, the music and the sound effects.

If harmony is achieved in these components, success will be held in the advertisement wedge. Unlike the jingle, a wedge transmits the message without appealing to the vocal musical interpretation of it. On TV, it is equivalent to the spot which has the same simplicity and generally the same standard

duration (30 seconds) only with the proviso that in the spot the image against audio is held and therefore another process is added into the recording. Wedges with low intention require grave and soft voices. Wedges with high intentionality require voices with high power and timbre. Equally to the jingles, wedges have commercial goals and therefore they adjust to tariffs according to duration (besides on-line duration). Nevertheless, there are wedges (or spots) which have neither commercial nor institutional goals.

5.4 Sensoriality: Stimulation, excitation, and sensation of the advertise message

Throughout time, technology has simulated real spaces which may convince the listener about a recreated situation which may imitate a real environment of living.

The ambiophonics system may convert the listening field into a scenario centered on the observer, almost tri-dimensional and clearly multi-sensorial.

5.4.1 The stimulation

When messages are constructed in monophonic or stereophonic systems, due to their spatial limitation in representation, it is not possible to present the messages in the omnidirectional concept that our auditory sensory system allows. This is one of the main advantages of the Ambiophonics system, as it has been mentioned on several occasions: the possibility to construct a spatial stimulation without directional limitations at a horizontal level.

5.4.2 The excitation

Every sensorial or physiological excitation implies an alteration of its receptor organs. These stimuli which excite the auditory organs obtain an automatic response into a physiological level: ear-drum contractions and dilatations, transmission through the middle ear bones and the process generated inside the internal ear, which yields into signals inside the central nervous system. After the stimulation, the excited organ recovers its precedent state. The excitation is, therefore, a passing modification follows by a return to the previous equilibrium. However, the excitation level is not associated with the communicative effectiveness level.

Therefore, the auditory excitation degree produced by the advertisement message should not encounter parallelisms in terms of greater or less efficacy, because this does not depend on the intensity degree in dB which applies the stimulus; a message which contains different sound intensity variables may place the emotional and meaning load into the fewer intensity passages which correspond to less excitation.

There is a tendency to believe that the greater intensity stimuli are the most effective ones. A more intense stimulus rises greater excitation, however, this does not mean that the message is more effective.

5.4.3 The perception

The perception is a psychophysical process by which the individual transforms the diverse sensorial impressions in a known object. It involves a process of four successive changes, which go from a more direct reaction into a more elaborated one: sensorial stimulation, sensorial excitation, sensation or emotional reaction and the hue.

Perception may be defined as the set of processes and activities related to the stimulation which the senses reach, through which, an information regarding the habitat, the actions and the inner state of the human beings is reached. It involves attention, interest, and memory.

5.4.4 The sensation

The sensation may be defined as the impression due to the direct action of the stimuli to the sense organs, or any internal impression, feeling or affection. The sensitive system has the ability to convert the physical and chemical stimuli into nervous impulses. Every advertisement has a solid emotional load, which may be associated to an affective link, but without losing the rational appearance. It is

understood that the emotional aspect may persuade more than the rational. It finds less barriers to achieve the message penetration and by general rule, it obtains more perdurability of it. Therefore, a situation drawn from a natural context as the ambiophonics system –able to simulate a global environment- has more potential to suggest than conventional stereo technique.

REFERENCES

1 Blauert, J., *Spatial Hearing: the psychophysics of human sound localization,* Verlag, MIT, (1997).

2 Espitia, x., Barco, F., Herrera M. Ambisonics, *Proceedings of the 24th International Congress on Sound and Vibration*, London, UK, 10–14 July, (2017).

3 Herrera, M., Calderon, L. Grubb Statistics, *Proceedings of the 23th International Congress on Sound and Vibration*, Athenas, Greece, 10–14 July, (2016).

4 Nicol, R., Emerit, M. 3D Sound reproduction over an extensive listening area: A hybrid method derived from holophony and ambisonic, *AES 16th International Conference.*

5 Glasgal, R. Ambiophonics 2nd edition. 2001.

6 Herrera, M. Compresor perceptual basado en la transformada Wavelet-Daubechies Tipo IV, con dos niveles de descomposición, *Congreso Iberoamericano de Acústica*, *FIA,* Valdivia, Chile, 2014.

7 Bosi, M., Goldberg R. E., *Introduction to Digital Audio Coding and Standards.* Springer, (2003).

8 Lord Rayleigh O. M. Pres. R. S. XII. On our perception of sound direction, *Journal of Philosophical Magazine Series 6,* **13** (74), (1907).

9 Macpherson E. A., Middlebrooks J. C. Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited, *The Journal of the Acoustical Society of America,* **111** (5), (2002).

10 Gardner, W. G., *3-D Audio using Loudspeakers.* Kluwer Academic Publishers, (1998).

11 Zhou, C., Hu, R., Tu, W., Wang, X., Gao, L. Binaural Moving Sound Source Localization by joint Estimation of ITD and ILD. *AES 130th International Conference (2011).*

12 Cha, H-T., Ryu, I-H, Seo, B-K. Implementation of 3D sound using grouped HRTF. *AES 34th International Conference: New Trends in Audio for Mobile and Handheld Devices (2008).*

13 Satongar, D., Pike, C, Lam, Y. W., Tew, A. I., The influence of headphones on the localization of external loudspeakers sources. *Journal of Audio Engineering Society*, **63** (10), 799-810, (2015).

14 Power, P., Davies, B., Hirst, J. Investigation of the impact of 3D surround systems on envelopment.. *AES 137th International Conference (2014).*

15 Glasgal, R., *Ambiophonics.* 2nd edition (2001).

16 Glasgal, R., *Surround Ambiophonics recording and reproduction.* (2003).