4 Spherical Microphone Probe for Surround Sound

Text and illustrations from an article by Leonardo Scopece published on Elettronica e Telecomunicazioni, April 2010

1. Introduction

The desire to feel *immersed* in an event, in the comfort of a living room, captured by spectacular images and surrounded by music and realistic sounds, has always represented a strong motivation to maximise the benefits of the steady progresses registered in the audio-visual and multimedia fields.

Since the 90s, users have been offered audio reproduction systems capable of ensuring listening sensations similar to those felt in a concert hall during a theatrical performance or a concert. *Home video* systems were first introduced in 1982, with the launch of Dolby Surround. The viewer feels right at the centre of the scene thanks to 5 loudspeakers (two of which are positioned behind the viewer), each fed by an audio channel.

Video quality levels available 30 years ago with *home video* systems were similar, in terms of definition, to levels available when the television was first introduced. Programmes were in colour, in 4:3 format, and broadcast (based on the PAL standard) or recorded on cassette (VHS). In fact, the first HDTV experiments started halfway through the 80s: 16:9 format images characterised by a number of elements (pixels) five times higher with respect to traditional TV standards. The video quality levels were equal to the ones which only in recent years have been made available to users. The first experiment of a digital format HDTV broadcast was carried out by Rai during the Italia ’90 FIFA World Championships. However, only twenty years later TV screen technologies (Plasma, LCD...), have made it possible for these new standards (in terms of both size and costs) to reach our homes, also thanks to the new DVD satellite and terrestrial broadcast technologies (and also based on progresses made in terms of the maximum exploitation of channel capacity).

During the last decade, technological progresses have increased their pace and the next step towards a full immersion in TV virtual reality will be made possible through the broadcasting of 3D programmes: the introduction of stereoscopic television, or 3D TV, appears to be imminent.

Apparently, such progresses are not equally fast and radical in the audio field.

Users can set up reproduction systems in their living rooms which are adequate both in terms of images and sound: large screens and surround audio systems. This is *home theatre:* a system which ensures the complete involvement of the viewer in the scene. However, quite often high definition images don’t go hand in hand with an equally spectacular audio quality.

It will be necessary to act along the entire production chain, starting from the first stage: production, video recording and post-production systems should all operate with high definition video signals, while audio recording and post-production systems should all be capable of coding surround signals.

Three years ago, the Rai Research Centre launched a vast experimental programme with a view to identifying the most effective systems for the production of surround audio programmes. The programme is aimed at setting up a system which will enable sound shooting and recording sessions with multi-channel systems using the existing infrastructures, with limited costs, in order to offer users a surround signal to complement the high-definition video signal, in its simplest form.

2. From experimentation to innovation

The previous chapters have described the stereophonic and holophonic technologies, and the experiments carried out during the 2007-2009 period.

Such experiments were carried out in different environments with different types of sound sources, using a microphone which has already been implemented by a number of North American broadcasters and tested by famous artists: the Holophone H2Pro.

The Holophone microphone is equipped with 8 capsules, seven of which are located around the peripheral area of the device and one inside, for low frequency collection. The 8 signals originated by the "head" are separated: the 8 channels can be directed to a mixer or coded as audio 5.1, 6.1, 7.1 or simply stereo.

The holophonic audio capture technology integrates the traditional sound capture technologies, which are characterised by specific characteristics of the reproduced sound:

- With the multi-microphone technology the sound is entirely present, focused, clean, but lacking depth, a "wall" of sounds originating from the instruments arranged on the stage.
- With stereophonic technologies, from MS to Stereosonic, AB or XY, sound starts acquiring a certain "depth".

New multi-channel technologies, such as the one implemented by Holophone, offer a more enveloping and natural sound, less "focused", but more realistic when compared to the sound obtainable with the technologies which have hitherto been used.

The results of the experimentation mentioned above are extremely interesting and have shown that the "ways of listening" to products obtained with multichannel sound shooting technologies vary.

We still have to assess the different perception by the audience due to the transition from the more artificial listening modes offered until now to a new way of listening, more enveloping and more real, although less “present”.

3. The system based on the H.O.A. microphone probe

Based on the acquired knowledge and the advantages of multichannel shootings, the Rai Research Centre has launched a project aimed at setting up an innovative system based on the Ambisonic technology (see Appendix).

The project has led to the development and patenting of an audio shooting system based on the HOA (High Order Ambisonics) Microphone probe by Rai and A.I.D.A., a spin-off of the University of Parma.

The system is aimed at shooting the sound scene and at providing three main results:

- Providing a multimicrophone or, alternatively, surround audio shooting;
- Enabling the microphone zoom function, i.e., change the directivity, both dynamically and in real time;
- Enabling the location of recording spots within the space both on the azimuth and median planes, again, dynamically and in real time.

3.1 The HOA probe and the Interface

The microphone probe (figure 1) consists of a sphere (diameter: 8.4 cm), the surface of which is covered by 32 high quality electret capsules.

The capsules ensure a 20 Hz-20 kHz frequency response, with a 32 mV/Pa sensitivity and a 135 dB dynamics before clipping.

The sphere contains the preamplifiers of the 32 capsules and the A/D converters.

The 32 digital signals are multiplied and transferred through an Ethernet cable (Class 5 or higher) measuring up to 140 m in length, to an interface (figure 2), from which they can be memorized on a USB device for subsequent processing or transferred to the processing unit.

3.2 Processing unit

The PU is the heart of the system. It ensures the processing of the audio signals, and the real-time complex calculations for the emulation of the filters which enable to synthesize the signals from seven virtual microphones.

Each of these virtual microphones has a latency time which is sufficiently low so as not to compromise the shooting, while ensuring good quality levels across the audio band base frequencies. Moreover, each microphone is synthesized based on the 32 component signals, captured by the individual capsules. The polar pattern may vary from omnidirectional to cardioid of various orders. The algorithms used are based on the Ambisonic principle: the sound information in a specific point in space obtained by detecting the pressure and speed of particles in such specific point. Currently, for each of the 7 synthetized microphones, the system enables the real-time processing of the information received from the 32 capsules and the synthesis of a virtual microphone Ambisonic equivalent up to the sixth order.

3.3 Pointing and the joystick

The system is aimed at maximising user-friendliness and ergonomics.

Fig. 1 - Up to the 6rd order Ambisonic

Fig. 2 - Audio Interface

Fig. 3 - Joystick

Fig. 4 – Displayed example of the polar diagram referring to two of the virtual microphones overlapping the image of the scene.

The user-friendliness in ensured by interfacing the processing system with a joystick (figure 3) and with a wide-angle video camera used to control the "pointing" of each individual virtual microphone.

Thanks to such devices it is possible to define in a simple way both the polar diagram and the pointing of each individual virtual microphone, independently.

The joystick enables to select each individual microphone, to dynamically "follow" actors on the scene and to set the directivity of a selected virtual microphone by displaying the corresponding beam with a coloured circle, the dimensions of which will intuitively describe its shooting angle specifications.

Information referring to the individual virtual microphones in use is displayed on a monitor, overlapping the video camera image (figure 4). It is therefore possible to decide in real time which polar diagram to use for each microphone, and the corresponding spatial positions.

If a scene is totally or substantially static, it might instead be useful to use a camera, in lieu of a video camera, to take a panoramic picture and memorize the positions of the actors on the scene without having to "chase" them, since they would be easily recalled, thus significantly simplifying the work (figure 5).

3.4 The control console

The system is managed through a control console (a notebook). It would also be possible to record on the notebook's hard disk, in real time, the 32 signals originated by the probe and/or the 7 synthesized signals corresponding to the virtual microphones.

The system management application enables the forwarding of the 7 system output signals to the listening space or to a mixer, through the ADAT port, or their recording on external audio recorder.

Naturally, the recorded signals may be later subjected to a post-production process. In particular, the signals corresponding to the 7 virtual microphones may be synthesized and modified during the post-production stage, starting from the complete 32 signals captured through the HOA probe.

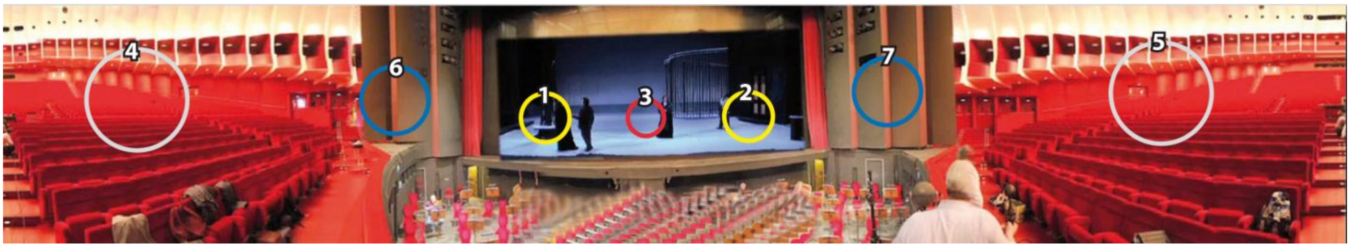

Fig. 5 – The choice of the polar diagrams and the selection of pointing features for each of the 7 virtual microphones can be carried out and memorized with the support of a panoramic photograph taken when the hall is empty or during the rehearsals. In the example provided, a sixth order microphone is oriented towards the centre of the stage (3), two 3rd order microphones (1 and 2) at +/-30°, two second order microphones (6 and 7) at +/-60° and two first order microphones (4 and 5) +/-130°. The chosen array may be modified during the recording or the post-production stages by selecting each individual microphone with the joystick.

4. Applications of the HOA probe-based system

The range of applications which may benefit from the use of this developed and patented system is extremely wide: we shall make a few examples referring to different audio shooting conditions. The experimentation stage involving potential users is still in progress, and is aimed at verifying the feasibility and flexibility of the system. It will thus be possible to find useful indications for the transition from the prototype stage to the production stage.

4.1 Orchestral performances

In the case of a symphonic orchestra, it is possible to simulate a multimicrophone audio shooting, for example by pointing 6 of the 7 virtual microphones towards specific points within the ensemble and dinamically controlling a seventh subject, a singer or a musician, who does not hold a static position on the stage.

Another example is represented by an audio shooting coded up to 7.0 by pointing 5 of the 7 virtual microphones frontally and the remaining 2 rearwards. Coded 5.1, 6.1 or 7.1 signals may be obtained by processing the 7 available signals and obtaining the LFE (Low Frequency Effect) ultralow frequency contribution, or using an additional microphone for the 20-110 Hz frequency range.

4.2 Sports events

In a football stadium, it is possible to obtain a surround effect by positioning the probe right by the field. It will thus be possible to capture also the sound of the ball-kicks and the impacts of the ball on the goal post or the crossbar.

During a cycling race, by capturing the sound of the crowd “running along” to support the racers, it would be possible to obtain a sound effect which would “come towards” the listeners and would then fade behind them.

4.3 Theatrical performances

As for orchestras, in a theatre it would also be possible to point 5 or 6 virtual microphones and to control the remaining microphones with one or two joysticks: this would enable a surround audio shooting coded between 5.0 to 7.0.

4.4 Television events

As regards a studio TV programme, it is possible to install one or two probes in a high-up position, and direct them downwards.

By “pointing” the virtual microphones corresponding to each probe with a fifth or sixth order directivity, it is possible to significantly reduce the background noise usually found in TV studios, such as the noise caused by air conditioning systems and lighting systems.

4.5 Radio events

In Radio studios, it would be possible to carry out spatial sound recordings that would make the listener more involved in what he/she is listening to, offering the listener feeling of being capable to “see” the position of the actors, speakers and musicians present in the studio, within a 360° space.

4.6 Talk-shows

The dynamic “following” made possible by the joystick would enable to follow, for example during a talk-show, a journalist moving towards the studio audience, who could be “pointed” in a static way by the remaining 6 microphones, also used for interviewing the guests.

By positioning the probe in a high-up position, over the audience, it would be possible to “point” the single components of the audience which will be interviewed with the joystick.

4.7 Sound scene shooting and reproduction of the result in another space

There might be limitations to the use of the system if the sound is reproduced within the same space: in fact, due to the high sensitivity of the probe microphones there is a high feedback probability (Larsen effect).

The system might instead find a wide range of applications for the shooting of the event and its real time reproduction in one or more spaces, also distant from the first one, ensuring a high ‘*presence effect’.*

4.8 Post-production

As mentioned above, audio data from the 32 capsules may be recorded and used for subsequent processing. This opens up an unprecedented range of post-production opportunities.

During the post-production stage it would be possible to modify the choices made during the recording stage. It would in fact be possible to reposition the virtual microphones within the set and redefine the directivity, in order to achieve the desired result.

The 7 virtual microphones limitation is referred to the current real time processing capacity and it might be increased in the future, while as of today it is possible to increase the number of virtual microphones during the post-production stage, naturally this precludes the possibility of operating in real time.

5. The next steps

The results achieved so far are extremely promising. Current experiments are carried out with a prototypal system: some of the components (interface, processing unit, console), might be integrated in order to enhance their portability and user-friendliness.

The testing of the various applications, carried out by involving the users in different environments and with various sound emission sources, is essential in order to identify a system which can be used for Radio-TV productions, the target of all research and development initiatives carried out by the Rai Research Centre.

The involvement of the industrial sector is instrumental to the development of a product which, based on the indications received so far, appears to be characterised by significant application and dissemination potential.

Appendix

The Ambisonic theory

Ambisonic is a method for recording, or sampling, and reproducing the information regarding a specific sound field.
In fact, the pressure and speed of air particles in a specific point in space are the information required to faithfully recreate the sound field. It is possible to exploit these two parameters in two main directions: synthesis of microphones or surround audio reproduction.

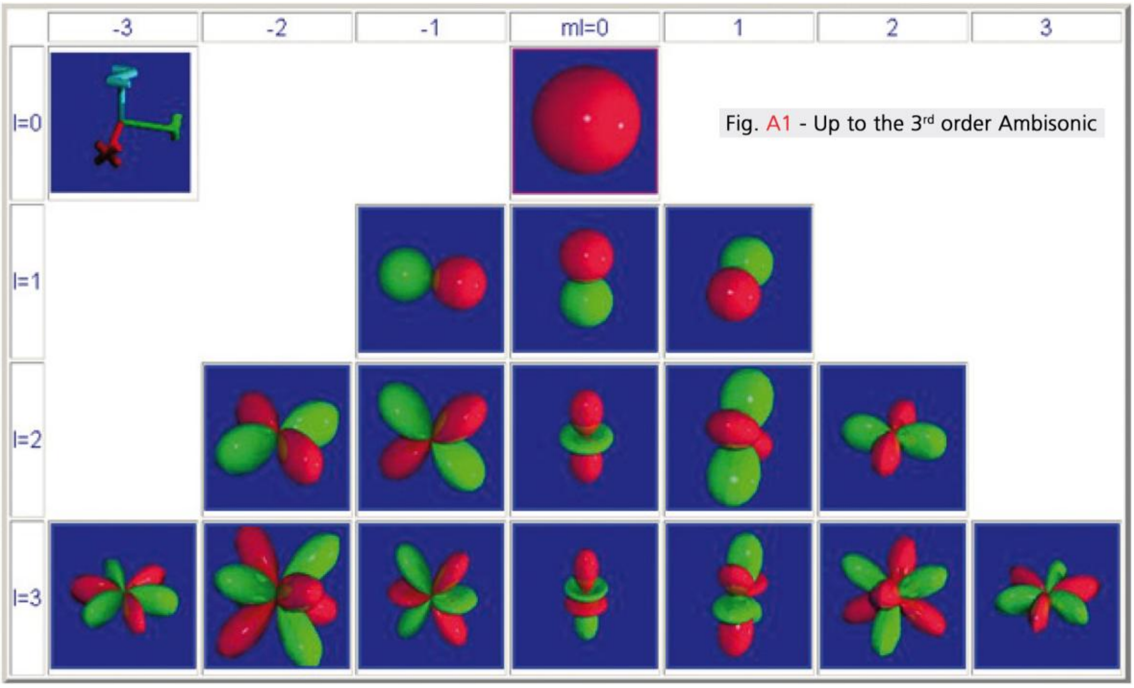

Fig. A1 - Up to the 3rd order Ambisonic

During the '70s, Michael Gerzon created the first probe capable of sampling pressure and speed in a specific point, relying on a tetrahedral array of microphone capsules with cardioid directivity.

Using three "figure-of-eight" microphones and one omnidirectional microphone, theoretically coinciding, we could consider sampling the order 0 and 1 harmonic spheres (directional functions): the omnidirectional signal referring to pressure data (W), the three remaining signals, oriented along the Cartesian axes, are proportional to the speed of air particles around the origin.

For each order "m" ambisonic there are (2m+1) components (figure A1).

Due to the impossibility of having 4 coinciding microphones, Gerzon thought of using 4 cardioid directivity capsules positioned on the faces of a tetrahedron. The conversion of the signals from the capsules (A-format) in order 1 Ambisonic signals (B-format) is carried out through a simple sequence of additions and subtractions, resulting in a series of signals virtually originated by other directions (figure A2):

W = C1+C2+C3+C4

X = C1+C2-C3-C4

Y = C1-C2+C3-C4

Z = C1-C2-C3+C4

Again, the capsules are not coinciding: the approximation used causes the misalignment of the capsules in terms of time/phase, leading to the “colouring” (filtering) of the B-format signal spectrum.

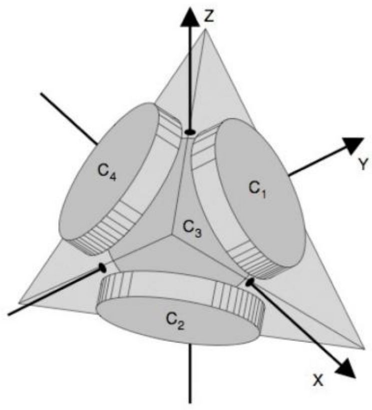

Fig. A2 – Tetrahedral microphone