

360 DEGREES VIDEO AND AUDIO RECORDING AND BROADCASTING EMPLOYING A PARABOLIC MIRROR CAMERA AND A SPHERICAL 32-CAPSULES MICROPHONE ARRAY

L. Scopece¹, A. Farina², A. Capra²

¹ RAI Research and Innovation Center, ITALY, ² University of Parma, ITALY

ABSTRACT

The paper describes the theory and the first operational results of a new multichannel recording system based on a 32-capsules spherical microphone array and some practical applications.

Up to 7 virtual microphones can be synthesized in real-time, choosing dynamically the directivity pattern (from standard cardioid to 6th-order ultra-directive) and the aiming. A graphical user's interface allows for moving the virtual microphones over a 360-degrees video image.

The system employs a novel mathematical theory for computing the matrix of massive FIR filters, which are convolved in real-time and with small latency thanks to a partitioned convolution processor.

INTRODUCTION

This research is the result of requirements emerged from the Centre for Research and Technological Innovation of RAI for a sound recording system capable of performing the same operations currently possible with a video camera, that is, focusing at a given point, panning, tilting and zooming.

While video has made great strides over the past decade, audio capture is still made with traditional mono or stereo techniques.

After 1990, when RAI performed experiments during the Soccer World Cup in Italy, research on high definition video stalled, mostly due to lack of technology that supports new ideas. TV sets could not any longer be those big "boxes" with CRT tubes that were used in the '90s, but had to evolve into what later became the flat screens we have today. But this had to wait for the industry to build plasma displays, LCD, OLED, LED, and all of this had a great development only after 2000, then having an exponential trend of technology with consequent reduction of video equipment's costs.

Where did the idea come from?

All this development regarding video had no equivalent in the audio field, where broadcasting is still done by means of mono, or at most two-channels stereo, shooting techniques. When about three years ago the RAI Research Centre has addressed this issue, the goal was not to seek for new technological solutions on the user side, as the user is already equipped with various home theatre systems, usually capable of 5.1 discrete surround, which are on the market since mid-90!

The purpose of RAI was to make multi-channel surround sound recordings employing as much as possible the existing infrastructure, at low costs and offering to the user the

easiest way to receive and play surround signal. In this way the audio scene may become as enveloping as the video scene, which is nowadays much more “immersive” thanks to new technologies in the field of television images, such as large-screen HDTV and Super High Vision, and now even stereoscopic 3D.

The research was focused on microphone systems, because the performance of currently available “discrete” microphone arrays do not offer the flexibility required in the audio-video production currently employed for broadcasting: although some of the discrete microphone arrays can be mounted directly over the cameras, so that pan-and-tilt are automatically synchronized with the camera movement, there is no way to zoom in the soundfield. Furthermore, the sound perspective should not generally be that resulting from a camera-mounted microphone array, as this would move completely the acoustic scene whenever the video image is transferred from a camera to another.

On the other side, the usage of a single, fixed microphone array results in a “static” sound capture, which do not correspond the dynamic video currently employed in broadcasting, where the camera never stay fixed for more than a few seconds.

Virtual microphones

In recent years a lot of research has been produced about technologies for capturing and reproducing the spatial properties of the sound. Most of the proposed approaches employ massive arrays of microphone and loudspeakers, and process the signals by means of very complex mathematical theories, based on various modifications of the classical Huygens principle.

These methods are based on mathematical representation of the sound field, which is decomposed in plane waves (Berkout et al. (1)), in spherical harmonics (Moreau et al. (2)) or in complex Hankel functions, (Fazi et al. (3)).

At playback stage, proper loudspeaker feed signals are derived, employing corresponding mathematical theories.

Whatever method is employed, at the end one can always think to the whole processing as the synthesis of a number of “virtual microphones”, each of them feeding a loudspeaker in the playback system.

We decided to NOT employ any mathematical representation of the sound field, searching for a numerical solution capable of providing directly the coefficients of the digital filters capable of synthesizing any prescribed virtual microphone, with arbitrary directivity pattern and aiming.

Albeit this approach can work, in principle, with any geometrical layout of the microphone array, we decided to develop our system around an high-quality 32-capsules spherical microphone array, recently made available on the market (mhAcoustics (4)).

The 32 signals are filtered employing a massive convolution processor, capable of real-time synthesis and steering of up to 7 virtual directive microphones, controlling their aiming and capture angle by means of a joystick or a mouse, and employing a wide-angle panoramic video camera and a graphical “view&point” interface for an easy-to-operate user's interface.

This can be done in real time and with small latency during a live broadcasting event; alternatively, the raw signals from the 32 capsules can be recorded, together with the panoramic video, for subsequent synthesis of the virtual microphone signals in post-production.

The virtual microphones being synthesized can be highly directive (with polar pattern constant with frequency, and with a beam width much sharper than a “shotgun”

microphone), and are intrinsically coincident, so the signals can be mixed without concerns of comb filtering; it is possible to continuously move their aiming for following actors or singers on scene, or for giving instantaneous miking to people in the audience.

Surround recording of a concert is just one of the possible scenarios for employing this approach, which has also been successfully tested for dramas, sport events, and TV shows in which there is systematic interaction of conductors and guests with in-studio audience.

A careful analysis of the performances of the new microphone system did show that frequency response, signal-to-noise ratio and rejection of off-beam sounds are better than those obtainable employing traditional processing algorithms applied to the same input signals, or dedicated top-grade ultra-directive microphones.

DESCRIPTION OF THE SYSTEM

Digital filters for virtual microphones

Given an array of transducers, a set of digital filters can be employed for creating the output signals. In our case the M signals coming from the capsules need to be converted in V signals yielding the desired virtual directive microphones: so we need a bank of $M \times V$ filters. As always, we prefer FIR filters.

Assuming x_m as the input signals of M microphones, y_v as the output signals of V virtual microphones and $h_{m,v}$ the matrix of filters, the processed signals can be expressed as:

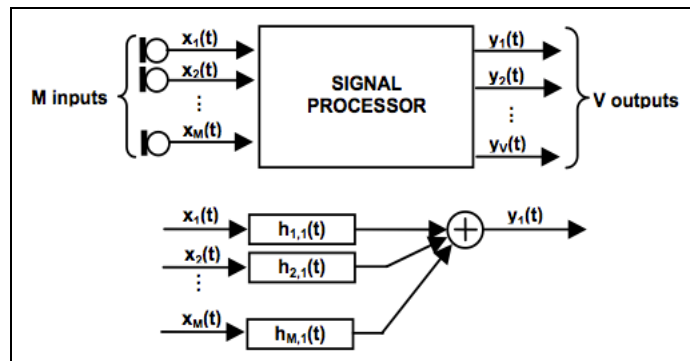


Figure 1 – Scheme of the signal processing

$$y_v(t) = \sum_{m=1}^M x_m(t) * h_{m,v}(t) \quad (1)$$

Where $*$ denotes convolution, and hence each virtual microphone signal is obtained summing the results of the convolutions of the M inputs with a set of M proper FIR filters.

In principle this allows for synthesizing virtual microphones having an arbitrary directivity pattern. In practice we decided, for now, to synthesize frequency-independent high-order cardioid virtual microphones, as shown in Figure 2, and to specify the aiming in polar coordinates (azimuth and elevation).

The processing filters h are usually computed following one of several complex mathematical theories, based on the solution of the wave equation [Berkout et al. (1), Moreau et al. (2), Fazi et al. (3)], often under certain simplifications, assuming that the microphones are ideal and identical.

In this novel approach no theory is assumed: the set of filters h are derived directly from a set of measurements, made inside an anechoic room. A matrix of measured impulse response coefficients c is formed and the matrix has to be numerically inverted (usually employing some approximate techniques, such as Least Squares plus regularization); in this way the outputs of the microphone array are maximally close to the ideal responses prescribed. This method also inherently corrects for transducer deviations and acoustical artefacts (shielding, diffraction, reflection, etc.).

The mathematical details for the computation of the processing filter coefficients are given in Capra et al. (5).

The microphone array

The experimentation described in this paper was realized using the Eigenmike™ microphone array produced by mhAcoustics (4).

As shown in Figure 3, The Eigenmike™ is a sphere of aluminium (the radius is 42 mm) with 32 high quality capsules placed on its surface; microphones, pre-amplifiers and A/D converters are packed inside the sphere and all the signals are delivered to the audio interface through a digital CAT-6 cable, employing the A-net Ethernet-based protocol.

The audio interface is an EMIB Firewire interface; being based on the TCAT DICE II chip, it works with any OS (Windows, OSX and even Linux through FFADO). It provides to the user two analogue headphones outputs, one ADAT output and the word clock ports for syncing with external hardware.

The systems is capable of recording 32 channels with 24 bit resolution, at sampling rates of 44.1 or 48 kHz.

The preamplifier's gain control is operated through MIDI control; we developed a GUI (in Python) for making easy to control the gain in real-time with no latency and no glitches.

Measurements of the microphone array were made employing the Exponential Sine Sweep (ESS) method, in order to obtain 32 Impulse Responses for each direction of arrival of the test signal.

These measurements were made inside an anechoic room, to avoid undesired reflections and to maximize the signal/noise ratio.

The loudspeaker and the anechoic room were kindly made available by Eighteen Sound, Reggio Emilia, Italy, who also provided the high-quality loudspeaker employed for measurements, as shown in Figure 4.

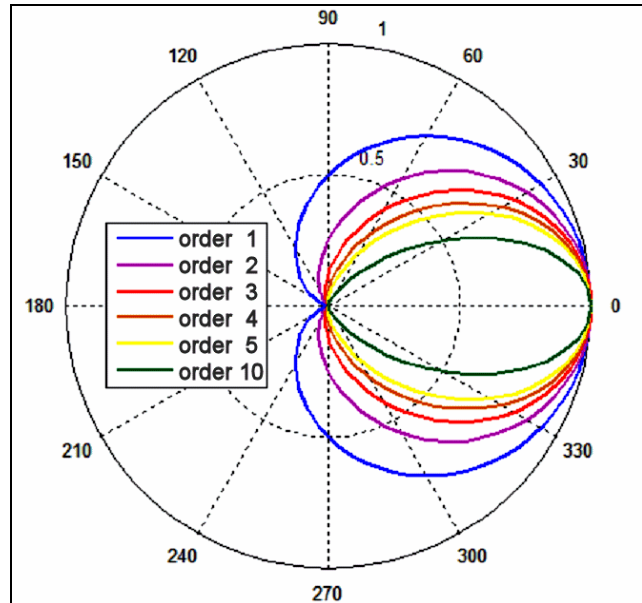


Figure 2 – Polar plot of virtual cardioid microphones of various orders (target patterns Q_n)



Figure 3 – The Eigenmike®

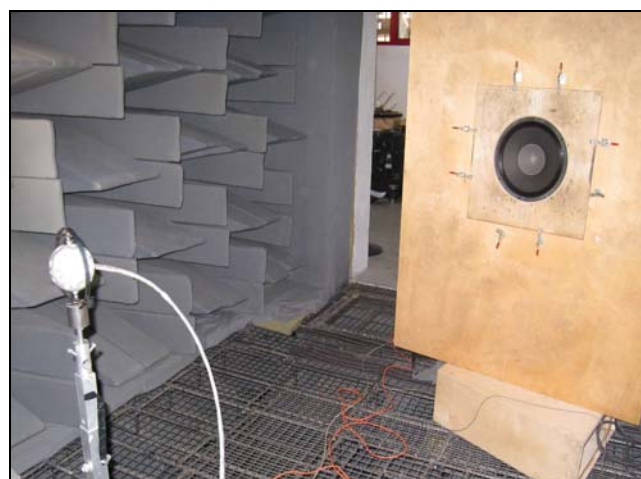


Figure 4 – The microphone array in the anechoic room

The array was rotated along azimuth (36 steps) and elevation (18 steps), using a movable fixture for the azimuth rotation and a turntable for the elevation. In this way we obtained 36 x 18 x 32 impulse responses, each 2048 samples long (at 48 kHz).

Synthesis and test of virtual microphones

In order to derive the matrix of filters, a Matlab script was produced.

The convolution of the FIRs matrix with the 32 signals coming from the capsules of the array should give as outputs the signals of virtual microphones with the desired characteristics. In the pictures below are shown some experimental results, showing some of the different directivity patterns obtained.

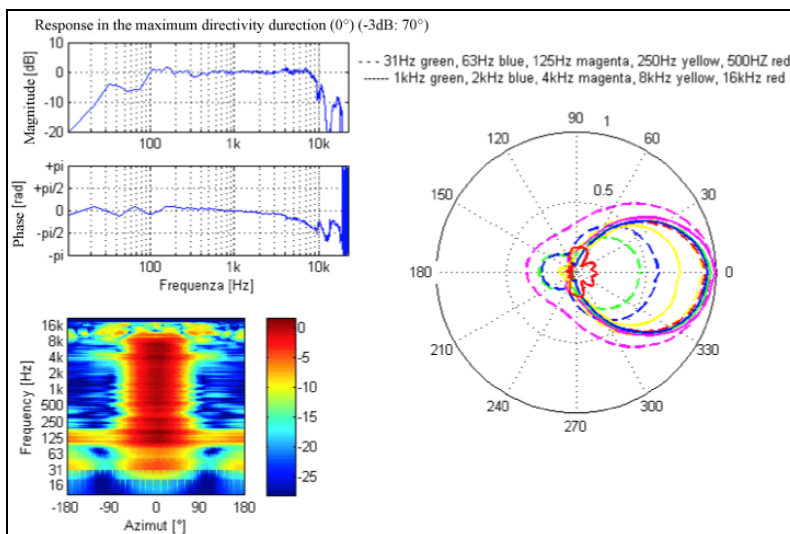


Figure 5 – 3rd order cardioid

Architecture and Graphical User's Interface

For using a system in broadcast production it should be very robust, without unnecessary complexities; on the other hand, we needed to provide massive computational power for performing all those FIR filters in real-time and with small latency, so we needed to add a dedicated "black-box" containing mini-ITX motherboard with a Quad Core processor.

All the signal processing is made by this computer. A laptop is used for visual control of the virtual microphone's characteristics over the IP network, a joystick/mouse is used for changing in real-time the directivities and orientations.

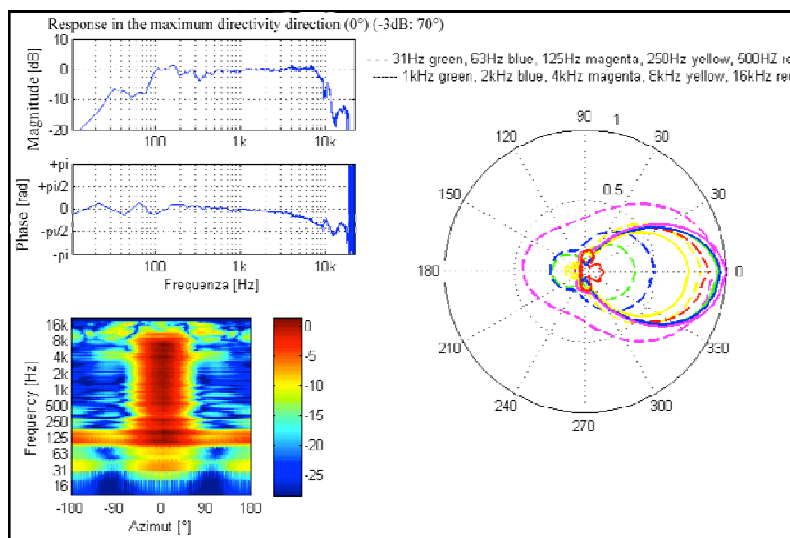


Figure 6 – 6th order cardioid

The scheme in Figure 8 represents the software architecture for the signal processing inside the “black box”.

The user interface is designed with the goal of following actors or moving sources, delivering live to the production the corresponding “dry” audio signals. For doing this, the GUI permits to focus the microphones on a video stream coming from a panoramic surveillance camera, equipped with a 360° parabolic mirror, placed close to the microphone probe.

This GUI is shown in Figure 9. A slider labelled "Transparency" is visible: it is useful to adjust the transparency of the coloured pointers overlaid on the live video image.

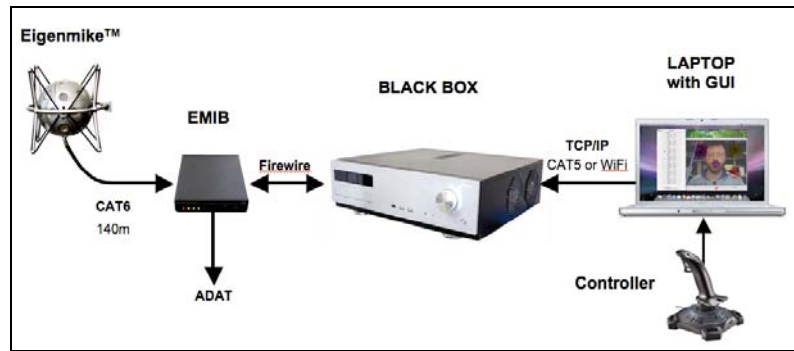


Figure 7 – Hardware architecture of the system

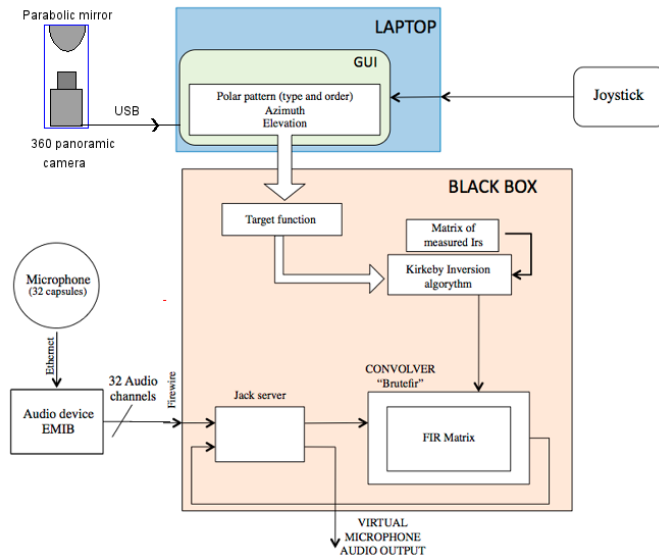


Figure 8 – Software architecture

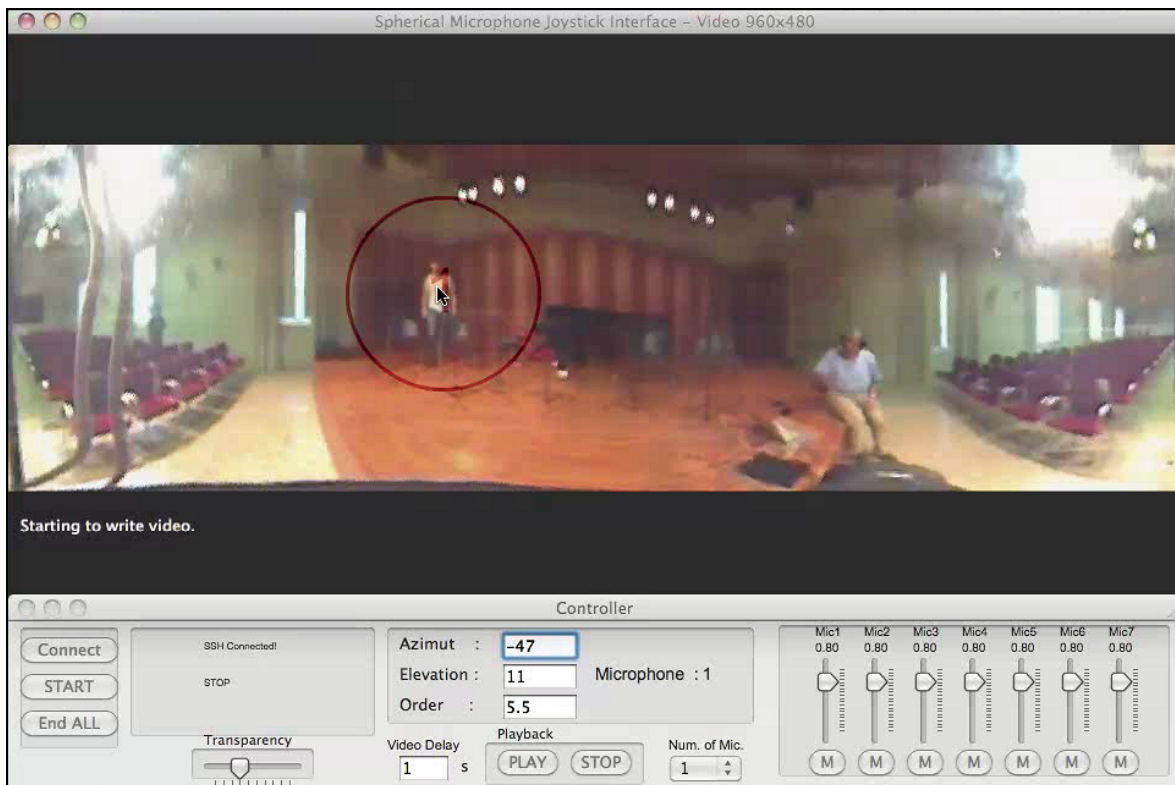


Figure 9 – GUI for realtime processing overlaid on a live video stream

APPLICATIONS OF THE H.O.A. (HIGH ORDER AMBISONICS) MICROPHONES

Examples of applications of the HOA Microphone can be various:

1. shooting an orchestra
2. shooting sports events
3. shooting theatrical events
4. shooting television events
5. shooting radio events
6. shooting a talk show
7. shooting a scene to play the result in another environment

Shooting an orchestra

By placing the probe in front of a symphony orchestra, just behind and above the director, you may want to target 6 of the 7 virtual microphone to fixed points over the stage, and dynamically control a seventh microphone, following a subject singing or playing music that is moving along the scene.

Another possibility is to fix 3-5 microphones to front and sides, and 2 towards the rear, to pick up "standard" surround formats ranging from 5.0 to 7.0. To get the additional ".1" channel (LFE) various types of bass management of the main channels can be performed, or a special omni-directional microphone can be added, particularly suited for the range of 20 to 110 Hz.

Shooting sports events

In a football (soccer) stadium, placing the probe on the sidelines, you can achieve a surround effect and simultaneously you might be able to capture the sound of the impact of the ball against the door frames or other obstacles, by "following" it with a continuously-moving virtual microphone.

It is also possible to "pick up" what players say, even if they speak very softly, employing a very directive virtual microphone aimed at their mouth.

During a bicycle race, or a running competition, taking the crowd with a virtual microphone that "runs" on the sides of the athletes, you can get a sound effect that "meets" and then "smoothes" at the back.

Shooting theatrical events

In theatre, you can shoot actors moving on stage, placing five microphones in a standard fixed arrangement and controlling the other 2 with 2 joysticks, following the actors. These "spot" mikes are then "mixed in" in the main surround soundtrack, ensuring that their voice is very crisp.

Shooting television events

In a television studio, you can put one or two probes suspended above the stage.

"Aiming" a virtual microphone with very high directivity to the mouth of every single participant, we exclude ambient noise normally present in the studio, like air conditioning and motorized lights. In practice, signals equivalent to wireless microphones can be recorded, without the need to give a wireless to every participant.

And if someone from the public decides to intervene, you can quickly steer a virtual microphone to him, instead of waiting that a valet brings him a wireless...

Shooting radio events

In radio studios the usage of several virtual microphones, instead of a number of real microphones given to actors and musicians, can provide a better mono-compatible broadcasting, avoiding comb filtering and interference between discrete microphones; in fact, these always capture also part of the sound of the other sources, with variable delays. The virtual microphones, thanks to their high directivity, maintain the presence and clarity that usually can be attained only by closely-miking each talker, singer or musician.

Shooting a talk-show

In a talk-show, placing the probe on top of the pit, the sound engineer controlling the virtual microphones can follow the journalist who moves in the studio going for example to its guests and speaking with them. Other virtual microphones can be kept pointed towards the public, awaiting for who wishes to intervene, and quickly "pointing" the joystick on who's speaking.

Shooting a scene to play the result in another environment

Due to the high sensitivity of the HOA microphone, it resulted risky to employ it for sound reinforcement (feeding the virtual microphone signals directly to loudspeakers placed in the same environment). This may cause Larsen effect, if a virtual microphone is erroneously pointed toward a loudspeaker.

However the HOA microphone system did show to work very well, also for real-time, live applications, whenever the sound captured by the virtual microphones is reproduced inside a different environment. This occurs, for example, in videoconferencing applications, or during TV shows when a remote connection with another place is performed

A further advantage of the HOA probe microphone is the possibility to select orientation and directivity of the virtual microphones later, during the post-production phase. During the event the whole set of 32 raw microphone signals is recorded on disk (the W64 file format is required in this case, for being able to exceed the limit of 2Gb, which is encountered very soon if recording in standard WAV or BWF formats). A panoramic 360-degrees video is also recorded, time-synched with the audio recording, employing the panoramic camera mounted close to the microphone.

During post-production, it's possible to reposition the virtual microphones in their place and redefine directivity, thanks to the panoramic video footage, which allows to follow the sound sources during their movement, as it was already described for live applications..

In a few words, it's like redefining the location and type of microphones after the event is recorded, something that was so far simply impossible.

In real time, no more than 7 virtual microphones can be processed, while during the post-production phase the number of virtual microphones can be increased to any number.

Experiments are being performed for checking the operational limits under various environmental and acoustical conditions, and different types of sound sources (human speech, musical instruments, sounds of the nature, vehicles, aircrafts, etc.).

The first, strong limit emerged till now regards the sensitivity to wind noise, as the HOA microphone probe is currently not equipped with a suitable windshield: a specific research project has been started regarding the construction of a windshield capable of reducing effectively wind noise, while maintaining proper cooling for the electronics incorporated inside the probe.

CONCLUSIONS

The goal of this research project was the realization of a new microphone system, capable of synthesizing a number of virtual microphones, dynamically changing their aiming and directivity.

The polar patterns of the virtual microphones can be varied continuously from standard types (i.e. omni or cardioid) up to very directive 6th order cardioids, which resulted narrower than top-grade “shotgun” mikes.

These virtual microphones are obtained processing the signals coming from a 32-capsules spherical microphone array, employing digital filters derived from a set of impulse response measurements. This provides several benefits:

- Widest possible frequency range for a given size of the probe, with low noise;
- Inherent correction for the dissimilarities between the transducers;
- Partial correction also for acoustic artefacts, such as shielding effects, diffractions and resonances.

The finalized system provides a revolutionary approach to sound capture in broadcasting and film/music productions.

REFERENCES

1. A.J. Berkhout, D. de Vries and P. Vogel, 1993. Acoustic control by wave field synthesis. *Journal of the Acoustic Society of America*, May 1993, 93(5) pp. 2764-2778.
2. S. Moreau, J.Daniel, S.Bertet, 2006. 3D sound field recording with high order ambisonics – objective measurements and validation of a 4th order spherical microphone. *Proceedings of 120th AES Convention*, Paris, France, May 20-23 2006.
3. F.M. Fazi, P.A. Nelson, 2007. The ill-conditioning problem in sound field reconstruction. *Proceedings of 123rd AES Convention*, New York, NY, USA, October 5-8 2007.
4. <http://www.mhacoustics.com>
5. A. Capra, L. Chiesi, A. Farina, L. Scopece, 2010. A spherical microphone array for synthesizing virtual directive microphones in live broadcasting and in postproduction. *Proceedings of 40th AES International Conference*, Spatial audio: sense of the sound of space, Tokyo, Japan, October 8-10 2010