5 3D - Virtual Microphone System High-directivity Microphone Probe

Text and illustrations from an article by Leonardo Scopece (Rai), Angelo Farina and Andrea Capra (University of Parma - Dipartimento di Ingegneria Industriale) published on Elettronica e Telecomunicazioni, April 2011.

1. Introduction

The previous chapter illustrates the research programme carried out by the Rai Research Centre and the Industrial Engineering Department of the University of Parma, which led to the design and implementation of an audio shooting and recording system based on the Ambisonic principle. The system uses an HOA probe and is known as 3D-VMS. The results achieved have enabled the definition of a patent registered by Rai and AIDA, a spinoff of the University of Parma.

The new system is based on a multi-capsule microphone probe aimed at maximising flexibility within audio/video productions in the broadcast field, reaching levels not obtainable with the "discreet" microphone models available on the market.

Discreet microphone models can be mounted directly on a video camera and their movement can automatically be synchronized to the camera movement, but they're not capable of zooming across the sound field.

The single fixed-position microphone provides a "static" audio shooting which does not correspond to video dynamics currently used in the broadcast sector, where images usually stay fixed only for a few seconds.

The 3D-VMS system enables a multi-microphone or, alternatively, surround audio shooting with a real-time dynamic microphone zoom function, and allows for positioning numerous audio shooting spots on the azimuth and median planes (up to a maximum of 7 virtual microphones).

This chapter represents an update of the previous one, which is referred to the 2010 activities, and provides further details regarding the basic system operating features and a number of considerations on experiments carried out in different environments and application areas.

2. Virtual microphones

In recent years, numerous research initiatives have been launched with the purpose of recording and reproducing the spatial properties of sound. Numerous approaches are based on the use of vast arrays of microphones and loudspeakers to process signals based on complex mathematical theories, deriving from the classic Huygens principle.

These methods are based on mathematical representations of the sound field which is subdivided into flat waves [1], spherical harmonics [2] or complex Hankel functions [3].

Irrespective of the methods used, the results may be calculated as the synthesis of a number of **virtual microphones**, each feeding a loudspeaker within the audio reproduction system.

The chosen approach does **not** use a mathematical representation of the sound field, while it is instead aimed at identifying a numerical solution capable of directly obtaining the filter coefficients used to synthetize each virtual microphone with arbitrary directivity and pointing values.

Although this approach is expected to prove effective, in principle, with any type of microphone array geometry, the system was developed using a spherical microphone probe consisting of 32 high-quality capsules, which has recently become available on the market.

The 32 signals are filtered using a powerful convolution processor capable of synthesizing up to 7 virtual microphones in real time and of defining their position and directivity using a joystick or a mouse as a control device. The management of the system by the user is facilitated by the use of a panoramic wide-angle video camera and an intuitive graphic interface.

The management can be carried out in real time and with low latency time during a live event. Alternatively, the “raw” signals from the 32 capsules may be recorded, together with the panoramic video, thus enabling their processing and the synthesis of virtual microphones during the post-production stage.

The synthesized virtual microphones may be alternatively directional (with constant polar diagrams with reference to frequency, or with a highly defined capture beam angle, similar to that of a rifle microphone), and are intrinsically coinciding, enabling the mixing of the signals eliminating comb-filtering issues; the microphones may be constantly moved on the stage to follow actors and/or singers, or to interview the audience.

The recording of concerts is only one of the possible contexts in which this scheme may be utilized, in fact, the system has also been tested for theatrical performances.

A careful analysis of the performances of the new microphone system has shown that the frequency response, the signal/noise ratio and the sound rejection levels outside the capture beam are better than those obtained using traditional processing algorithms applied to the same input signals, or dedicated and highly directive microphones.

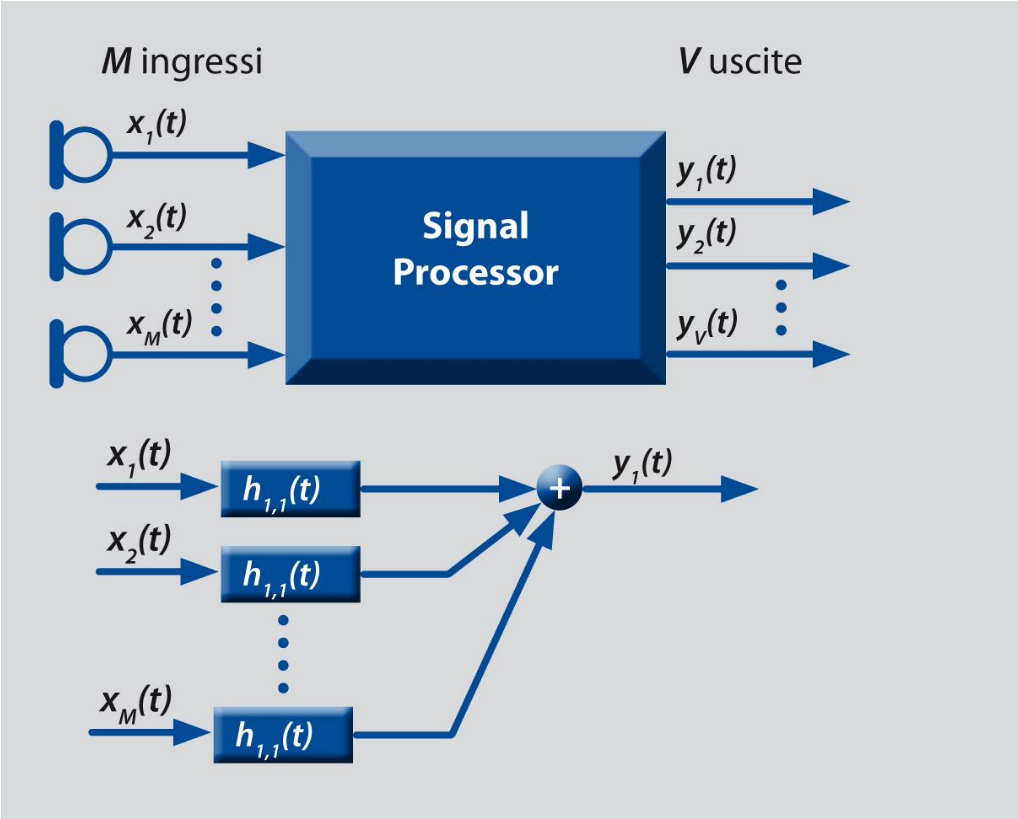

Fig. 1 – Signal processing scheme.

3. System description

3.1 Digital filters for virtual microphones

Given an array of transducers, it is possible to use a set of digital filters to create the output signals (figure 1). In this specific case it is necessary to convert the ***M*** signals originated by the capsules into ***V*** signals corresponding to the desired virtual microphones: It will be necessary to use an ***M*x*V*** matrix of FIR filters.

Given $\boldsymbol{x_m}$ as the input signals from *M* microphones, $\boldsymbol{y_v}$ as the output signals from *V* virtual microphones and $\boldsymbol{h_{m,v}}$ as the filter matrix, the processed signals may be expressed as:

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

Where * represents the convolution, i.e. each signal from the virtual microphone is obtained by adding the results of the convolutions of the ***M*** inputs with an adequate set of ***M***

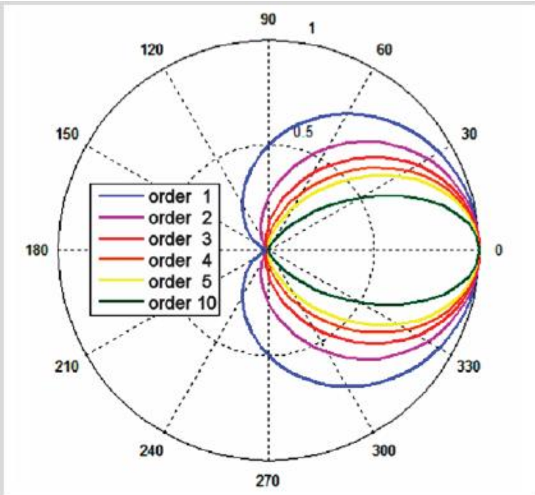

Fig. 2 – Polar diagram of virtual cardioid microphones of various orders.

In principle, this approach allows to synthetize virtual microphones characterized by an arbitrary directivity pattern. In practice, it was temporarily decided to synthetize virtual cardioid microphones of higher orders, irrespective of the frequency, as shown in Figure 2, and to specify the pointing direction in polar coordinates (azimuth and elevation).

The *h* filter coefficients are usually calculated following one of the complex mathematical theories based on the solution of the wave equation [1,2,3], often subject to some simplifications: it

FIR filters.

is assumed that the microphones are ideal and identical.

Fig. 3 – Microphone probe.

The planning criteria, instead, are not based on a specific theory: the *h* filter sets are directly derived from sets of measurements carried out in an anechoic chamber.

A matrix of impulse response coefficients based on the measurements is then created; the inverse matrix is calculated numerically (implementing approximation methods such as the least squares plus regularization method), this ensures that the microphone array outputs are normally very close to ideal responses.

This method also provides for the correction of the deviations of the transducer and acoustic artefacts (masking, diffraction, reflexion etc.). The mathematical details for the calculation of the filter coefficients are illustrated in [4].

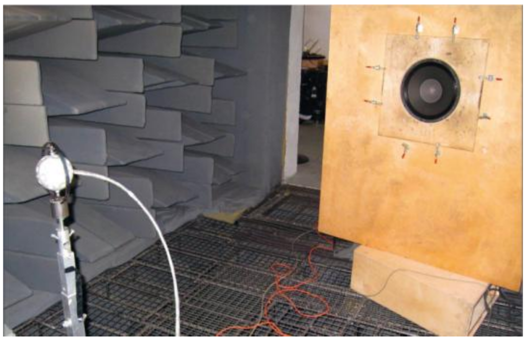

Fig. 4 – The microphone system in the anechoic chamber.

3.2 The microphone model

The experiment described in this article was carried out using the Eigenmike™ microphone array produced by MH Acoustics [5].

This microphone probe (figure 3) consists of an aluminium sphere (radius: 42 mm) with 32 high-quality capsules distributed over its surface. The microphones, preamplified and converted from analog into digital are inside the sphere and all the signals are transferred to the audio interface using a CAT-6 cable and an A.net based Ethernet protocol.

The audio interface is called EMIB and is based on the TCAT DICE II chip. It operates with Windows, OSX and Linux operating systems using FFADO. The board is equipped with two analog headphone outputs, a digital ADAT output and a word clock port for synchronization with external equipment.

The system is capable of recording 32 channels with a 24-bit resolution and a 44.1 or 48 kHz sampling frequency.

Gain control of the preamplifiers is carried out through MIDI control. A python language GUI was also developed, to enable simplified gain control without any latency or interruption.

3.3 Experimental characterization of the model

The microphone array measurements were carried out using the ESS method, to obtain 32 impulse responses for each direction of the test signal.

The measurements were carried out inside an anechoic chamber, in order to avoid unwanted reflections and to maximise the signal/noise ratio.

The loudspeaker and the anechoic chamber were provided courtesy of Eighteen Sound (Reggio Emilia), which also provided a high-quality loudspeaker for the measurements, as shown in Figure 4.

The system has been rotated along the azimuth (36 steps) and in elevation (18 steps) using a mobile azimuthal rotation device and a rotating platform for the elevation. This has provided 36x18x32 impulse responses, each consisting of 2,048 samples (48 kHz).

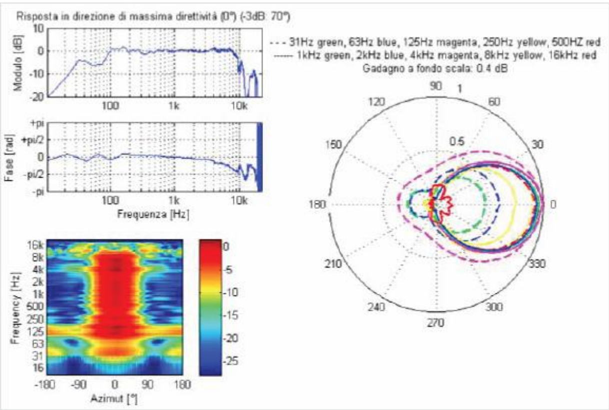

Fig. 5 – 3rd order Cardioid

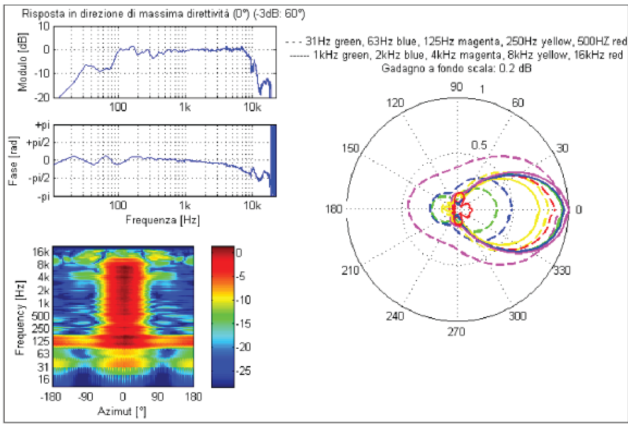

Fig. 6 – 6th order Cardioid

3.4 Synthesis and testing of the virtual microphones

The filter matrix was achieved using a MATLAB script.

The convolution of the FIR matrix with the 32 signals originated by the array capsules must provide the virtual microphones signals as an output, with the desired characteristics. Figures 5 and 6 illustrate testing results showing some of the different directivity profiles obtained.

3.5 Architecture and Graphical User Interface

In order to be used in broadcast productions, a system needs to be highly "robust", without any unnecessary complication. It is necessary to rely on a high processing capacity in order to generate and use the FIR filters in real time, with a low latency time. To this aim, the system is equipped with a dedicated signal processing unit containing an extremely powerful mini-ITX motherboard with a Quad Core processor. The audio signals are processed by this unit.

A notebook connected to the processing unit by means of an IP network is used for the visual control of the characteristics of each virtual microphone. For the real-time modification of the directivity and orientation of the virtual microphones, the user will rely on a joystick or a mouse (figure 7).

The scheme in Figure 8 illustrates the software architecture of the signal processing unit. The user interface is designed to follow the moving actors or sound sources, allowing the real-time validation of the audio signals position. To this aim, GUI (figure 9) allows to focus the microphones on a video flow starting from a panoramic surveillance video camera located near the microphone source.

Fig. 7 – System Hardware components. a Microphone probe, b EMIB interface, c Signal processing unit, d control console (notebook), e pointing device (joystick); f CAT 6 cable, in this application it can operate up to a length of 140 m, g ADAT optical interface, h firewire interface, i Ethernet CAT 5 cable or WiFi connection, l cable or wireless connection.

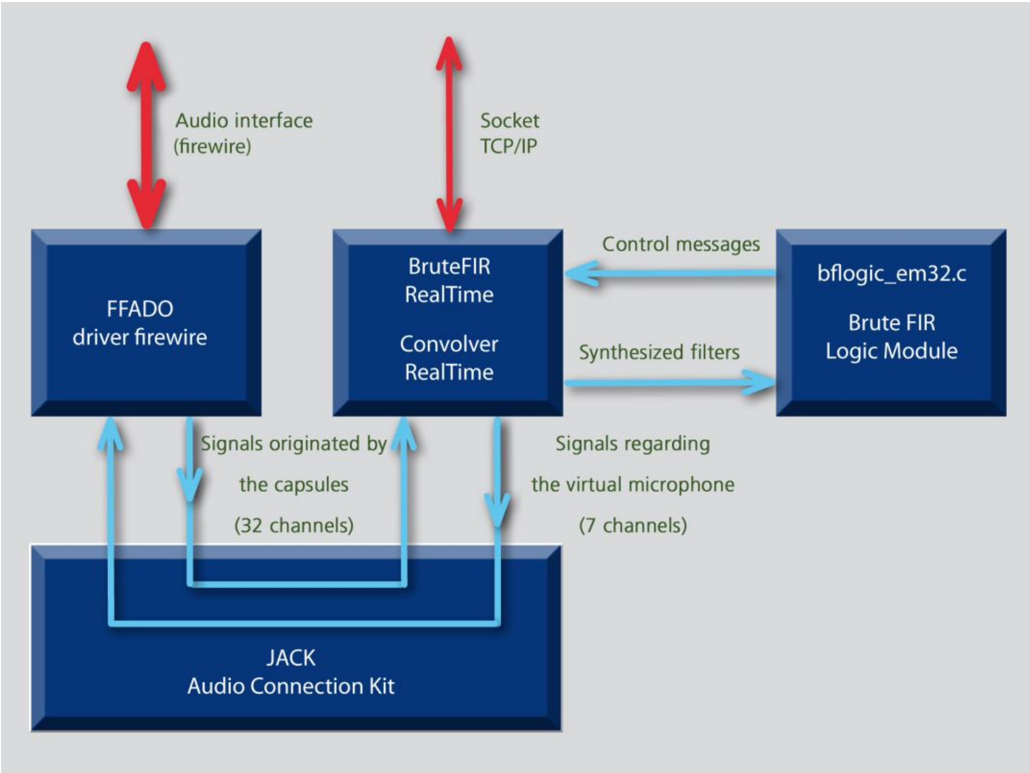

Fig. 8 – Software architecture

Fig. 9 – GUI. Please note the slider used to fine-tune the transparency of the coloured pointers overlapping the video image.

Fig. 10 – Turin, May 13, 2010. Classical music concert at the Rai Auditorium.

4. Application examples and experiments

The previous chapter describes the events to which the system may advantageously be applied: orchestral performances, sports events, theatrical performances, TV or radio events, talk shows and reproduction of an audio scene in an environment different from that where the audio shooting is carried out.

During recent months, numerous experiments have been carried out, all aimed at verifying the characteristics of the system in various shooting, production and post-production contexts. In some cases the experiments provided for a comparison between the holophonic system (Holophone H2Pro microphone), and the 3D-VMS system.

4.1 Concert held at the Rai Auditorium

A test regarding the audio shooting of a concert was carried out at the Rai Auditorium in Turin (figure 10). The test, aimed at comparing an audio shooting with a holophonic microphone and with the 3D-VMS system, took place during the rehearsals for a Sostakovich concert.

The two types of microphones were placed on two poles, located in close proximity to one another.

The result, with the microphone positioned behind the conductor, was extremely good, present and pleasing for both systems.

The recording carried out with the holophonic microphone is more “vivid”, but this does not necessarily imply a more pleasing result in comparison to the 3D-VMS system. At times, the holophonic result registered some non-coinciding signals, which proved to be absent with the 3D VMS system.

4.2 Concert held at radio studios

The audio shooting took place during the memorial concert for the victims of the earthquake in L’Aquila, the “Per non dimenticare” event, in a Rai Radio Studio.

Also in this case (figure 11) the microphone probe was in a central position, behind the Conductor. A video camera to interface the 3D-VMS system and the holophonic microphone was installed on the same support.

The audio results were satisfying, also thanks to the acoustics of the studio and the presence of a live audience, which acted as a good absorber.

The result of the comparison with the holophonic microphone is entirely subjective: the holophonic signal appeared less “dark” when compared to the ambisonic signal, but in general it was less satisfying.

Fig. 11 – Rome, April 14, 2010. Classical Music Concert at the Studio Radiofonico Rai in Via Asiago.

Fig. 12 – Turin, September 9, 2010. Melevisione TV Programme in a TV Studio at the Rai Production Centre.

4.3 Programme in a TV Studio

The experiment carried out during an episode of the Melevisione TV programme (figure 12) yielded results of no great significance because the probe had to be positioned at a distance to avoid the disruption of the works, in fact the scenography of the format takes up most of the studio space. In spite of the unfavourable conditions, however, the sound of the voices was good, although at times it was “out of focus”.

4.4 Opera: “La Bohème” at the Teatro Regio

The 3D-VMS system was used during the rehearsals (in front of an audience) at the Teatro Regio in Turin (figure 13).

Fig. 13 – Turin, May 21, 2010. The “La Bohème” rehearsals at the Teatro Regio.

The microphone probe, for direction purposes, was placed on a pole in front of the first row of seats, not in a central position, but slightly moved to the right (approx. 20°) from the centre, looking towards the stage, with the orchestra pit located between the probe and the proscenium.

The experiment was aimed at verifying the possibility of "shifting" the 7 virtual microphones towards the centre during the post-production stage, while maintaining unchanged the spatial ratio between them.

From the acoustic point of view, the result was not positive, because the voices of the singers were distant and severely "disturbed" by the music coming from the pit. It was only during the solos that the voices were captured in a cleaner way, although they still felt too "distant".

4.5 Theatre: "Arlecchino" at the Piccolo Teatro

Unfortunately, during the audio shooting of the play "Arlecchino servitore di due padroni" at the Piccolo Teatro in Milan, the probe could only be installed in a very inadequate position (figure 14): in the upper circle, in a lateral position, with a minimum view of the stage, and with a span of approximately 10°.

Fig. 14 – Milan, October 20, 2010. The Piccolo Teatro stages the play "Arlecchino, servitore di due padroni", shot in 3D HDTV for the video and 3D-VMS for the audio. The panoramic picture shows the microphone probe, in an evidently inadequate position.

The recovery of the sound data required more than two weeks of intense post-production work. The result went beyond expectations, it was in fact possible to “extract” the dialogues of the numerous actors, in spite of the critical operating conditions (the actors were turning their backs to the microphone, in presence of a strong murmur in the background).

4.6 Concert at the Conservatorio Giuseppe Verdi Hall

During the Concerto in D major op.35 for violin and orchestra by Tchaikovsky (figure 15), the pole holding the microphone probe was positioned at the centre and behind the conductor.

Fig. 15 – Turin, November 22, 2010. Concert at the Conservatorio Giuseppe Verdi Hall.

The result obtained was extremely satisfying and it required very little post-production work.

The best recording results are achieved when it is possible to position the microphone in the optimal frontal position, where the disturbing “audience effect” is minimised.

4.7 Concert in a TV studio

“Mise en abime” is a composition by Maestro Yuval Avital, written for a mass instrumental and choral ensemble, which uses space as an essential element of the orchestral score. Rai, in close collaboration with the Franco

Parenti Theatre, Maga Global Arts and the University of Milan, has tested the 3D-VMS system for the audio shooting of this extremely peculiar composition.

The experimental audio shooting was carried out at the Rai Production Centre Studio 3 in Milan (figure 16).

Fig. 16 – Milan. February 27-28 and March 1, 2011. Performance of the "Mise en abime" composition at the Rai Production Centre, Studio 3. The composer, on the left, analyses the desired audio shooting features using the 3D-VMS system.

The musical ensemble, arranged in a circular order, is particularly complex, and consists of 32 accordions, 2 bass accordions, 2 bass tubas, 1 bass clarinet, 1 soprano, 1 percussion and 1 didgeridoo. The choir consists of over 100 non-professionals arranged around an external circle and executing an alphabetical and graphic score. The ensemble is directed by four conductors, two for the choir and two for the instruments.

The microphone probe was positioned at the centre of the two circles. Furthermore, to ensure further audio support, 8 extra cardioid microphones were positioned, to cover an ideal subdivision of the stage in 8 different sectors.

The test produced a good quality quadriphonic product, which was the targeted result.

4.8 Opera: “Nabucco” in Rome, Montecitorio

Unfortunately, the audio shooting conditions for the performance of “Nabucco” at the Chamber of Deputies in Montecitorio, in Rome (figure 17) by the Rome Opera Orchestra directed by Maestro Riccardo Muti, were not ideal.

In fact it wasn’t possible to position the microphone probe at the centre of the Orchestra, as it was initially envisaged, because it would have obstructed the TV images. Consequently, the probe had to be positioned in the circle, approximately 15 metres from the orchestra, behind the position of Maestro Muti.

This limitation, together with the imperfect acoustic of the Chamber of Deputies generated a disappointing result, except for the full orchestra and choir passages.

Fig. 17– Rome, March 22, 2011. Opera: “Nabucco” at the Chamber of Deputies.

4.9 Opera: “Rigoletto” at the Teatro Regio

The test based on the performance of “Rigoletto” at the Teatro Regio (figure 18) was aimed at evaluating a more complex application of the 3D-VMS system.

Fig. 18 – Turin, April 9, 2011. Opera: “Rigoletto” at the Teatro Regio.

The test envisaged the use of a single microphone probe located at the beginning of the proscenium, in an almost central position, near the prompter’s booth, in order to obtain the recording of the 32 channels. During the post-production stage 7 frontal microphones, aimed at the singers, were synthetized and, based on the same multi-channel shooting, a further 7 microphones aimed in the opposite direction, towards the orchestra, were synthetized. The results have not been analysed yet, however, the expectations are good, based on what was perceived in the headphones during the shooting.

Fig. 19 – Turin, April 20, 2011. Opera: “La Traviata” at the Teatro Regio.

4.10 Opera: “La Traviata” at the Teatro Regio

With regard to “La Traviata” (figure 19) two microphone probes were used: one in the same position used for “Rigoletto”, and the second on a stand inside the pit, behind the Conductor. The results are still awaiting an analysis, but we expect that the configuration will yield a more defined reproduction of the audio components, i.e. the singers and the orchestra. It will be possible to evaluate the possible signal problems due to the mixing of the recordings from the two probes and the consequent implications for the final product.

5. Conclusions

The project was aimed at designing a microphone system capable of synthetizing a significant number of virtual microphones, for which it was possible to modify position and directivity features, both dinamically and in real time. The newly developed system enables the definition of virtual microphones, the polar diagrams of which may differ from the standard types (such as omnidirectional or cardioid), all the way up to highly directive 6th order cardioids, a type of directivity which has proven to be more narrow when compared to “rifle” microphones.

The project methodology offers significant advantages:

- ✓ The possibility of operating within a wide range of frequencies, for a specific probe size;
- ✓ The correction of the capsules structural differences;
- ✓ The partial correction of acoustic artefacts, such as masking effects, diffractions and resonances.

Particular attention was devoted to the operating characteristics, specific to the envisaged applications:

- ✓ The microphone probe was adapted to broadcast specifications: the colour does not reflect the stage lights; it is equipped with a handle to make it similar to a conventional microphone; it operates at the professional standard frequency of 48 kHz; the user interface (GUI) is intuitive.
- ✓ The unification of the signal processing and the control unit (used for recording purposes and for the user interface) is currently in progress; processing will also be entrusted to the notebook (the current control unit) thanks to the higher CPU processing capacity.
- ✓ This will facilitate the synchronization of the service video shooting (required for pointing the virtual microphones) and the time-code, which is essential for the production and post-production stages.

The system offers a radically new approach to audio shooting in the broadcasting field and for the production of films and music, and the envisaged innovations will soon overcome all current limitations, which are typical of an experimental product, and will thus ensure significant improvements from the operating, technical and dimensional points of view. The wide-ranging tests have shown that thanks to the operating flexibility the system may find advantageous applications within both the production and post-production stages.

Acknowledgements

For their contribution to audio shootings, editing and technical and musical consultancy services we wish to thank our Rai colleagues: Enrico Cavallini and Salvatore Cangialosi from the Research Centre, our colleagues from the Technological Strategies-Technical Quality department and the TV Production Centres in Milan and Turin.

In particular, we wish to thank the colleagues of the Rai Radio Services in Turin, Vatican Radio and the Audio/Video Technical Department of the Chamber of Deputies in Montecitorio, Rome.

Bibliography

1. A. J. Berkhout, D. de Vries, and P. Vogel, "Acoustic control by wave field synthesis", Journal of AES, 93(5):2764-2778, May 1993.
2. S. Moreau, J. Daniel, S. Bertet, "3D sound field recording with High Order Ambisonics - objective measurements and validation of a 4th order spherical microphone", 120th AES Convention, Paris, France – May 20-23, 2006.
3. F. M. Fazi, P. A. Nelson, "The ill-conditioning problem in Sound Field Reconstruction", 123rd AES Convention, New York, NY, USA – Oct. 5-8, 2007.
4. A. Capra, L. Chiesi, A. Farina, L. Scopece, "A Spherical Microphone Array for Synthesizing Virtual Directive Microphones in Live Broadcasting and in Post Production", 40th AES Int. Conf.: Spatial Audio, Tokyo, Japan, 8-10 Oct. 2010.
5. http://www.mhacoustics.com