Immersive audio at Roland Garros

LIDWINE HÔ, FRANCE TÉLÉVISIONS, **LEONARDO SCOPECCE**, RAI AND **ALBERTO CIPRIAN**, RAI

The 2015 Roland Garros French Open international tournament took place between 23 May and 07 June in Paris, as tennis fans will not need reminding.

The producer of the international TV signal, France Télévisions, and in particular, Francetvsport, its production unit (Tillière Production), its digital entity (FTV Éditions Numériques - FTVEN) and its R&D department (Innovations & Développements), covered seven tennis courts using five Outside Broadcast (OB) vans to produce the international HD video and 5.1 audio signals that were distributed worldwide.

Inside the Roland Garros Stadium was the RGLAB booth, created by FTVEN. This was dedicated to innovative technologies and it allowed the public to discover 360°video, live UHDTV with high dynamic range (HDR) pictures and binaural immersive sound.

During the two weeks of this event the Innovation & Développements team also experimented with binaural acquision and an experimental 3D virtual microphone provided by our colleagues in the RAI. UHDTV pictures were produced by a locked-off camera on the main court during the championship; its signals were broadcast by DTT, satellite and IP for live reception and display on three UHD screens in the RGLAB booth, together with headphones to listen to the binaural audio coming from the main court.

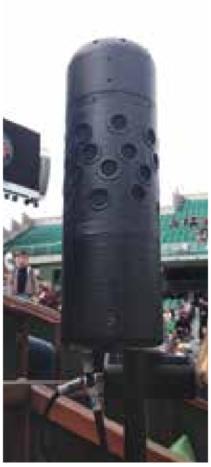
In addition, on 05 and 06 June, three UHD cameras were used to make a live UHD multicamera production from the main court.

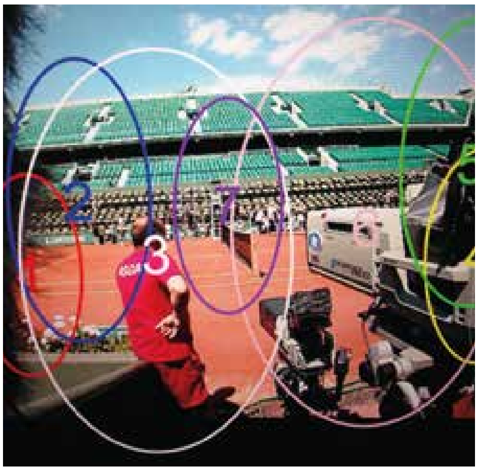
Binaural audio

In the context of 4K video shooting, where wide shots and slow transitions are important elements of the production grammar, native binaural miking was the simplest and most appropriate solution. In placing a binaural (dummy) head above each UHD camera, the listening and visual axes were in permanent alignment and provided a naturally immersive soundscape for the viewers equipped with headphones in the RGLAB.

Using this arrangement, the movements and distance of the tennis player were readily perceived by the viewer and the realism of the on-court voices, together with the emotions and reactions of the spectators were extremely impressive.

Electrically, each binaural, dummy head contains two microphones fitted in its 'ear canals'. The stereo signal produced was plugged directly into the camera microphone inputs; the binaural sound was then carried alongside the UHD video produced by the cameras.





On the two days of multicamera live UHD production, the camera switching on the video mixer generated a General Purpose Interface (GPI) command that was sent to the audio mixer to enable it to follow the on-air camera (the audio mixer faders were slaved to the GPI). It turned out that the major audio issues concerned the adjustment of the decay, sustain and release times associated with the audio switching; something that needed to be done manually most of the time.

It may have been a simple technique, but it allowed a strikingly immersive live soundscape to be created at minimal expense of time and technology.

Then we added 'close miked' sounds provided by the RAI's 3D Virtual Microphone System (VMS) to enhance the soundscape.

The Virtual Microphone System

The 3D VMS device was developed and patented jointly by the University of Parma and the RAI CRIT (its Turin R&D centre). Physically it consists of a single probe fitted with 32 microphone capsules in a spherical or cylindrical array together with a processor to create a single output containing elements of all 32 microphone signals that travel down a single Ethernet cable to its complementary processing unit in the audio control room. Acquisition with this system is more akin to multichannel shooting than miking or mixing. Up to seven virtual microphones can be synthesized in real-time, but many more can be derived in post-production processing. The azimuth, elevation and directivity characteristics of each virtual microphone can be independently and freely selected in real time using the software in the control room processor that applies digital filtering to each of the 32 capsules in the 3D VMS.

The power of this technique lies in the fact that the coefficients of the digital filters are computed from experimentally measured data rather than using the theoretical computation used in Ambisonics techniques that also employ microphone arrays. The experimental approach provides a better integration of the actual mechanical characteristics (small differences in the performance of the transducers and diffractions and reflections imposed by the body of the microphone array) of the microphone array than anything that is theoretically computed.

Additionally, there is a 360° camera in the 3D VMS (whose video signal is also conveyed in the single Ethernet cable) which aids in visualising the placement of the 7 virtual microphones on a screen.

Software modules are being developed for use with the 3D VMS. One of these performs automated tracking of actors or singers in a scene using a face recognition engine. The spatial coordinates of the person are then computed and are used to drive the position of each virtual microphone appropriately. Another software module aims at optimizing the spatial coverage of the sound along the three (X,Y, Z) axes so that a top view and a side view (corresponding to camera positions) may be generated in addition to the front view that is present.

Experimenting with the 3D VMS at Roland Garros was very informative and it points the way to new approaches to acquiring immersive audio.

The future is object-based audio (OBA)

In addition to the three binaural stereo signals and the outputs of the 3D VMS there were all the outputs of the numerous conventional microphones that are usually rigged throughout the tennis court by the international signal producer. All these audio resources were separately recorded in the OB van and they will be used later for OBA applications and experiments.