OBJECT- BASED AUDIO FOR TELEVISON PRODUCTION

Steven A. Silva

21st Century Fox, USA

ABSTRACT

Object-Based Audio (OBA) is the next exciting breakthrough in television production. It will provide personalization and an enhanced listening experience as revolutionary as sound was to motion pictures. The age of personalization has arrived and TV consumers can view any program, at any time and on virtually any media device.

The next generation of audio encoders will have the capabilities to create OBA in television production and post-production. The beneficial features of OBA include audio personalization for language selection, dialogue enhancements and options for the hearing impaired. OBA provides the viewer with the ability to customize their viewing for any type of program in any viewing setting.

The future audio codec technologies will enable audio production mixers, producers and broadcasters to produce customized audio for the viewer. This process begins at the original mix location and continues through the broadcast chain to delivery on any consumer device. These encoders will have the capability to emit surround sound and immersive sound with over 100 channels with objects either separately or in combination with each other. Scene based audio will also be a feature of the next generation codecs enabling the mixer to represent the sound image instead of channels.

This paper will describe the evolution from current channel based TV production to the next generation of multi-featured audio encoders with OBA and the potential benefits they will offer to all types of TV viewers.

INTRODUCTION

Since the beginning of TV production, channel based audio mixing has been the norm. Based on the creative vision of the director, the first audio mixers configured their mixing boards for microphones placed at critical locations within the stage or attached to the actors to capture their voices. Early TV shows comprised of drama, comedy, musical and variety genres as shown in Figures 1 & 2. For a musical show various microphones were

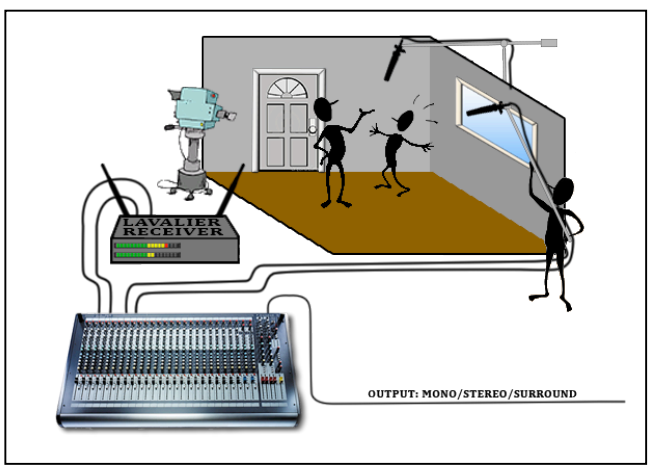

Figure 1 – Stage Drama Configuration

placed amongst the musicians and singers. Drama, variety and comedy shows would have a boom microphone, stage microphone or lavaliere microphone to pick up the talents voice.

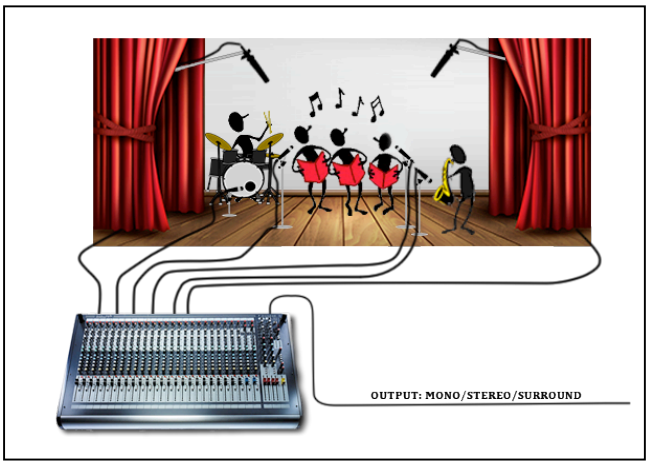

Figure 2 – Musical / Variety Configuration

As sporting events were included in TV broadcasts, a new challenge arose. The events action occurs on a field or court and is not precisely choreographed as with a drama type show. The solution for this setting was to have microphone placements established by a general game template, with announcers at different locations and additional sound sources from multiple locations as illustrated in Figure 3.

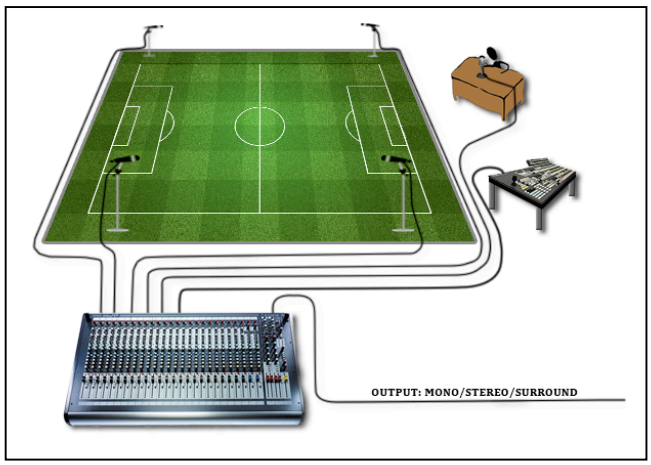

Figure 3 – Stadium Configuration

Once the microphone routes are established, all of the different microphone tracks are mixed together and coupled into a final mix. The final channel mix will be mono, stereo or surround sound. The master channel mix prohibits the independence of the singular microphone tracks for broadcast distribution. The individual tracks cannot be uncoupled from this final mix and personalization is unavailable.

All TV genres captured through channel based mixing create a remarkable life-like representation of the directors' vision. However, the viewer's only personalization/interaction is to control the loudness and the dynamic ranges available at the viewing location with the volume/loudness control.

In a live TV broadcast, the viewer hears the event as the sound mixer is creating the show. In a non-live TV show, the original mix is generally sent to a post-production facility for audio sweetening. These sweetened shows are generally drama, documentary and variety shows where immediate broadcast is not demanded. The sweetening process may involve the addition of objects along with its process for enhancing the original mix. The sweetened version is then sent to the broadcaster for a scheduled broadcast and possible archival at the broadcaster's facility. The next generation of audio encoders will allow the addition of personalization tools with OBA that will enhance the viewer's experience past channel-based limitations.

OBJECT-BASED AUDIO MIXING

Currently, all TV shows are captured in a channel based medium. In recent years there have been some theatrical releases incorporating OBA. They were produced with the next generation of audio encoders not yet employed by TV broadcasters. Presently, with the multitude of playback devices driving consumer demand for personalization, the channel based paradigm is rapidly becoming outdated. The Internet Age with second screen

devices generates opportunities and demands for OBA programming. With OBA on the horizon, there will be a need to broadcast this new medium to numerous hardware setups (i.e. Smart TVs, tablets, computers and smart phones) without the requirement for multiple and additional delivery networks.

In OBA mixing, objects are created in the audio track from a microphone source or another pre-produced audio source. These object-based tracks have accompanying metadata, permitting the track to be kept discrete with its unique arrangement details. Unlike with present codecs, the object can be custom mixed into a channel bed allowing multiple options currently unavailable to the producer and broadcaster. Ultimately, these options will provide the consumer with a number of personalized programming scenarios.

Boundaries constrain the mixer in a channel based world. The mixers resourcefulness is limited during the final mix-down to channels due to the inability to access objects. One feature of objects would enable the mixer to use vectors instead of panning for accurate sound positioning. This feature would mimic the action of the sports venue or the drama occurring on the TV stage with special audio effects. If the mixer was to create an objects only mix, the individual tracks could be kept unique and re-mixed in any configuration. The consumer could receive these individual tracks and create personal mixes for any media device ranging from their home theatre to any other mobile device.

The current surround sound broadcasted to consumers has six channels. Channel 1 is left, channel 2 is right, channel 3 is center (mainly for dialogue), channel 4 is for low frequency enhancement, channel 5 is left surround, and channel 6 is right surround. In surround mixes there is zero personalization for the viewer. The center channel, which is intended to be dialogue, will usually have other content from other tracks mixed into it or some other form of crosstalk from other channels. The channel mix allows the viewer of a sports show only one announcer, which is predominantly mixed in the center channel and is usually in one language.

A properly balanced 5.1 surround mix generally gives better intelligibility to the dialogue in the center channel and is clearer/cleaner than the stereo phantom center mix configuration. The viewer could personalize the mix by increasing the volume of the center channel of a surround mix with the desire to increase the intelligibility/comprehension of the dialogue. Unfortunately, with this approach, the results in quality will usually be poor and not be equivalent to an object-based program with dialogue personalization for volume and dynamic range.

Objects would enable the sound mixer and broadcaster more options to the final mix. The mixer at a sporting event could assign tracks for announcer objects in different languages; special sound effects to imitate the action of the players (e.g. skier swooshing about the course, or the sound of a baseball hitting a bat) or other audio sources from the event producers special effects library. In a drama show, using objects, the mixer could create added suspense by accurately mimicking the steps of an actor crossing a stage or an outside effect with greater accuracy, creating added suspense through objects. All of these audio objects are depicted with 3-dimensional coordinates and a mixture of other source qualities.

The broadcaster could customize the objects into a template at the mixers board or at different locations. The broadcaster would have the ability to insert objects for various dialogues, languages, commentaries, special effects and/or custom objects for the hearing impaired.

METADATA FOR OBJECT-BASED AUDIO (OBA)

The key component that makes OBA possible is the insertion of specialized object metadata. This insertion is done at the location of the original mixing board utilizing a next generation encoder-mixing tool. This mixing tool is generically called a "sidecar". The encoder manufacturer could have a sidecar mixing board/tool ideally located in the audio booth by the main mixing board. If the audio mixers location does not have enough space, the sidecar could be located elsewhere in the production center. The sidecar-mixing tool may require an additional audio mixer.

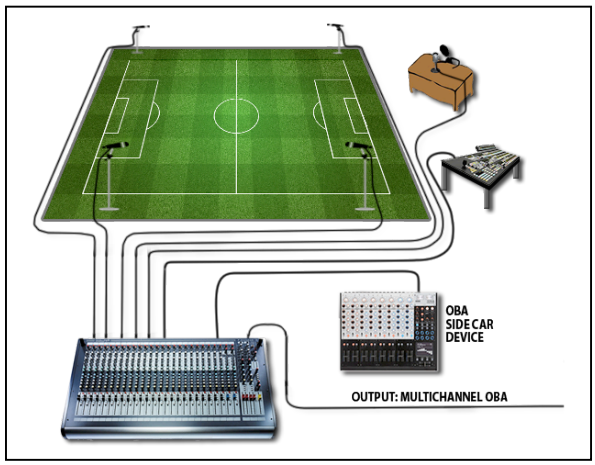

Figure 4 – OBA Stadium Configuration

The sidecar would tie into the main mixing board via a cable or Wi-Fi connection. The audio mixer and the sidecar mixer would determine what tracks would be objects. The sidecar mixer would author the metadata with possible added direction from the production team. Another option could be a plugin board/module in the main mixing console. This plugin board/module would duplicate the sidecar operation and remove the sidecar unit in the mixing area. Depending on the complexity of the event, a sidecar mixer may not be needed, leaving the authoring of the metadata to the audio mixer as shown in Figures 4-6.

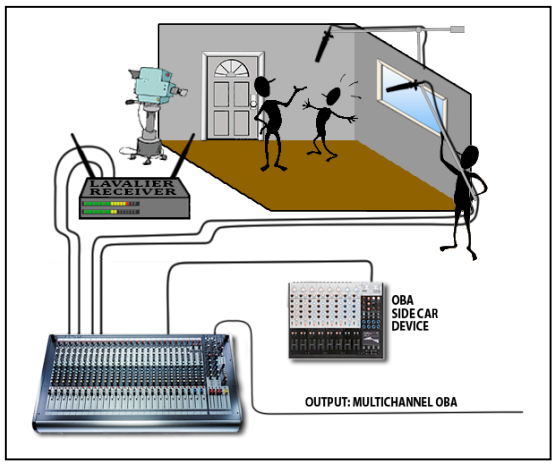

Figure 5 – OBA Stage Configuration

After the television show is completed, the integrated object-based mix can be sent to a post-production house. There it can be sweetened, further customized with various objects with descriptive metadata and then given to the broadcaster for emission. The post house and the broadcaster will have the ability to form a number of custom renderings for various broadcast formats. These renderings could have options for channel formats, objects with different languages, commentaries, loudness and dynamic range selections.

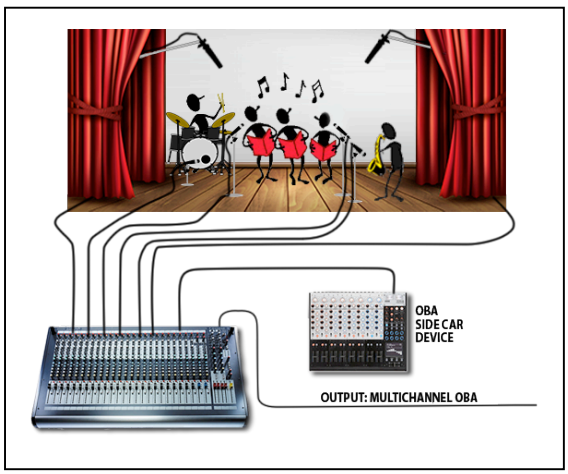

Figure 6 – OBA Variety Configuration

With a live TV show, the object-based mix is streamed to a broadcast center as it is being created. At the center, additional object metadata can be inserted, along with mix customizations for various renderings and then emitted from the broadcast facility.

The insertion of objects into the audio mixers creative tool set will remove the limits of channel mixing through the addition of objects

authored with metadata. The mixer will be able to author characteristics to the audio track simplifying and enhancing the audio process. The result will give the viewer personalization options at any location with any viewing device.

FUTURE BASED AUDIO ENCODERS

Manufacturers of the next generation of audio encoders have been displaying features at recent trade shows not available in the current generation of broadcast audio codecs. Current audio encoders are channel based with up to six (6) channels for surround sound.

One feature of future based encoders will be immersive sound with capabilities greater than 100 channels, being either discrete channels or objects or a combination of channels and objects. Immersive sound systems will create an immersive experience for the viewer from the addition of overhead speakers. At a minimum, it will be comprised of 12 channels, which go to eleven speakers and one low frequency speaker. Four of those speakers will be above the viewer's head to give immersion to the viewers' listening experience. Those channels could also be created as objects or a combination of both providing unlimited flexibility for the viewer. The immersive system could exceed the initial 12 channels and exceed 23 channels for home theatre systems.

Scene based audio creation will be another feature in the next generation of encoders. Its' design intent is to represent the action of a sports venue or stage drama as a field of sound pressure values at all points in that space over time. The goal of this feature is to replicate an absolute and true representation of the performance.

Current audio compression bit rates range from 32 Kbps to 640 Kbps. Major broadcasters generally compress the surround 5.1 mix to 384 Kbps for premium broadcast shows and use lesser bit rates for other platforms. The future audio encoder manufacturers claim that a combination of 12 channels and or objects along with metadata will be possible with less than 384 Kbps with equal or greater audio fidelity. These encoders compressed signal quality will allow multiple formats of lower bitrates for all types of smart devices and mobile phones. The encoder's output will be ASI and IP capable enabling flexibility for SDI and future IP systems.

OBA BENEFITS FOR THE HEARING IMPAIRED

Two groups of consumers who will greatly benefit from OBA are the hearing impaired and the aging population. Statistics from the World Health Organization state there are over 360 million people worldwide who have disabling hearing loss and impairments. Age demographics reveal that the population of hearing impaired will continue to increase.

One of the adverse effects of hearing loss is increased difficulty to discern dialogue surrounded by other channel sounds, channel crosstalk and background noise in a TV/Film mix. In a non-impaired hearing viewer, these channel mix characteristics are usually experienced with a minimum of complaints to producers and broadcasters. However, even a properly created channel based mix can often be difficult to understand for the hearing impaired.

In a channel based world there is no simple fix for this problem. Fortunately, with the addition of objects, loudness solutions for the hearing impaired will be available to enhance their viewing/listening experience and reduce complaints about unintelligible mixes to the broadcasters. Separate dialogue tracks where the volume can be increased for intelligibility with the option to lower the other channels will be a major benefit to these viewers.

OBA TV PRODUCTION

Television producers and broadcasters will approach OBA TV production with the “crawl, walk, run” philosophy. With this approach, the initial “crawl” period will emphasize critical production objects and be highly experimental. To optimize personalization, critical object selection for productions will be focused on presenting highly intelligible dialogue with multiple languages and commentary selections.

One excellent example of object-based production would be a racing car event where multiple cameras are tracking various drivers. In this production, viewers could select a driver and hear their conversation with their team captain at a desired volume and without any other channels as illustrated in Figure 7. The viewer would also be able to select commentaries about the race and the driver’s histories, etc. The viewer would also be able to select the ambience of the racetrack with the selected drivers’ dialogue at the desired loudness level. If a number of people are watching this event on a large monitor and each viewer has a second screen with headsets, the personalization can be customized for each viewer. In this scenario, each viewer could customize the audio via the second smart device for language, commentary or just listen to any dialogue at any volume level without interfering with the experience of other viewers.

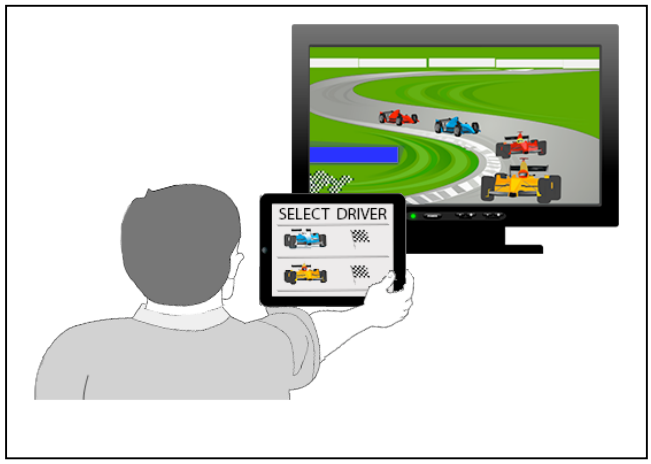

Figure 7 – OBA Second Screen Interplay

This experimental phase of sports production will include microphone trials using the plethora of microphone configurations and multi-microphones. The sound localization from OBA and the special effects created in a vector regime will provide more accuracy to the event. This will enable the viewer to choose certain dialogue options of the sporting event rather than listening to the general broadcast. Objects will enable a sports viewer to hear their local teams and not have to listen to the general broadcast as one would in a channel based broadcast.

It is foreseeable that sporting event producers will be the pioneers in OBA TV productions. They will have the opportunity to provide the personalization features that viewers have always wished for and change the way viewers will consume sporting events forever. Basic personalization features for sporting events will include multiple languages, announcers from the teams/contestants, event commentaries and custom event objects for each sport.

Programming templates for live and non-live productions will be created collaboratively amongst the broadcasters, producers and production groups. They will define the channel and object options broadcasted to the viewers and will become standardized. These standardized templates will allow the TV production community to maintain quality control throughout the creative chain to broadcast emission resulting in an unparalleled personalized viewer experience.

CONCLUSIONS

With UHD-TV and higher video resolution TV sets already in the consumer market place, it will only be a brief period of time before OBA and immersive sound will be joining these video formats via the next generation of audio encoders. This audio/video combination will provide the viewer with a brilliant visual and sound experience personalized to their viewing desires and personal needs. OBA and immersive sound will create innovative commercial ventures for programmers, broadcasters and multichannel video programming distributors (MVPDs).

REFERENCES

1. Riedmiller, Jeffrey C. and Tsingos, Nicolas, Dolby Labs Inc. 2015, Recent Advancements in Audio – How a Paradigm Shift in Audio Spatial Representation & Delivery Will Change the Future of Consumer Audio Experiences, Spring Technical Forum Proceedings.

2. Shirley, Ben and Oldfield, Rob, 2015, Clean Audio for TV broadcast: An Object-Based Approach for Hearing-Impaired Viewers, Journal of the Audio Engineering Society, Vol. 63, No. 4, April.

3. Bleidt, Robert, Borsum, Arne, Fuchs, Harald and Weiss, S. Merrill, 2014, Object-Based Audio: Opportunities for Improved Listening Experience and Increased Listener Involvement, SMPTE Journal.

4. Ruiz, Alejandro Gasull, Sladeczek, Christoph and Sporer, Thomas, 2015, A Description of an Object-Based Audio Workflow for Media Productions, AES 57th International Conference.

5. Oldfield, Robert, Shirley, Ben and Spille, Jens, 2014, An object-based audio system for interactive broadcasting, AES Convention.

6. Melchior, Frank and Spors, Sascha, 2010, Spatial Audio Reproduction: From Theory to Production, AES Convention.

7. Claypool, Brian, Van Baelen, Wilfried, Van Daele, Bert, 2012, Auro 11.1 versus object-based sound in 3D: All Aspects Compared, Barco.

8. Riedmiller, J., Mehta, S., Tsingos, N., Boon, P., 2014, Immersive & Personalized Audio: A Practical System for Enabling Interchange, Distribution and Delivery of Next Generation Audio Experiences, SMPTE Annual Technical Conference,

ILLUSTRATIONS

All Figure illustrations created by Ed Essl.

ACKNOWLEDGEMENTS

I have been very fortunate to have been involved in many aspects of the TV/Film and Broadcast industry. I would like to acknowledge the individuals who have played a pivotal role in my career: Andrea Berry, Steven Berry, Daniel R. Brewer, Todd Daly, Jim DeFilippis, Mel Duff, Richard Friedel, John Palacio Sr., Robert Reynolds, Anthony Rinaldi, and Clyde Smith Jr.