Audio Engineering Society

Convention e-Brief

Presented at the 145th Convention
2018 October 17–20, New York, NY, USA

Stationary Music from Users' Viewpoint in VR Applications

Sungsoo Kim, Sripathi Sridhar

New York University, 35 W. 4th St, New York, NY 10012

Correspondence should be addressed to Sungsoo Kim (sk6740@nyu.edu)

ABSTRACT

The ultimate goal in virtual reality (VR) is to achieve complete immersion in terms of audio and video, where background music is typically included to keep users absorbed in a game or 360-video content. This paper explores a multichannel loudspeaker configuration to anchor the background music to the user's viewpoint in VR. To that end, an evenly-spaced octagonal loudspeaker configuration is implemented in order to anchor the background music using head tracking data. The real-time panning is achieved through Vector-Base Amplitude Panning (VBAP). This paper also describes a demo interface built using the Oculus Rift in Unity and Max/MSP, as proof of concept.

1 Introduction

The ultimate goal in virtual reality (VR) is to achieve complete immersion in terms of both 360° audio and video. Accurate localization of sound sources is one of the most important components in VR as it ensures a more natural immersion for users. Hence, headphone technologies including head-related impulse responses (HRIRs) for binaural rendering, and Ambisonics, have been the subject of active interest in the VR development community. On the other hand, this paper explores possibilities of using multichannel loudspeaker configurations in VR. Multi-channel systems provide a great sense of space, and could reduce ear fatigue typically associated with extended headphone use. Such systems for VR can be used in demo installations and home theater-scale setups, both of which can provide professional VR experiences to the user.

In audio post-production of VR games or 360-video content, background music is typically included to keep users absorbed in the game content. When reproducing VR sounds in a multi-channel configuration, sounds that are not in view may be used to trigger the user's head movement. For example, a car driving behind users causes them to turn around, or take a particular action in the context of the game. Hence, VR developers can take advantage of sound localization to direct the user's attention, so as to deliver artistic intentions effectively. Thus, certain important sound cues can be moved around the user based on the user's head movement. Meanwhile, background music does not need to be controlled based on the head's movement in an artistic aspect, unless the VR content is solely focused on immersive sound recordings.

As the first step of a study on audio reproduction in multichannel loudspeaker layouts for VR, the authors implemented an evenly-spaced octagonal loudspeaker configuration that applies real-time panning in order to anchor the background music to the user's viewpoint.

Interactive panning over a multichannel loudspeaker layout can be achieved through several means such as Vector-Base Amplitude Panning (VBAP), which allows a sound source to be positioned anywhere in the space between two loudspeakers [1]. Sound objects in a multichannel reproduction can also be positioned by other methods such as Distance-Based Panning (DBP), which uses all the available

speakers to position a source instead of a pair at a time [2].

2 Background

2.1 Vector-Base Amplitude Panning

Vector-Base Amplitude Panning (VBAP) is a method that allows a virtual source to be positioned anywhere along the arc between two loudspeakers. Based on the desired position of the virtual source, the gain values of the loudspeakers can be calculated.

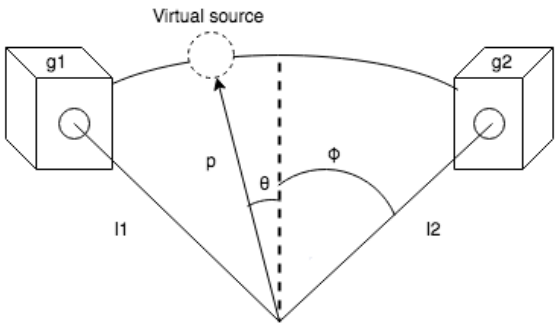

Figure 1. Vector-Base Amplitude Panning

In Figure 1, l_1 is the vector depicting the position for speaker 1, l_2 depicts speaker 2's position and p that of the virtual source. θ and φ represent the virtual source and speaker angles respectively. As described in detail by Pulkki in [1], the position vector of the virtual sound source is given by

$$p = r \, . \, \theta, \tag{1}$$

where radius r is set to 1.

$$P = G \, . \, L_{12} \tag{2}$$

based on the intensity panning law, where L_{12} is the speaker position vector matrix, P is the position vector, and G is the gain vector. Thus,

$$G = P \, . \, L_{12}^{-1} \tag{3}$$

gives us the required gain values for the two speakers which are used to create the virtual sound source. Using two gain factors for the loudspeaker pair, final channel outputs can be obtained by multiplying the gain factors with the corresponding signal data.

This process is repeated for all the virtual sources and loudspeaker pairs, giving the real-time loudspeaker gain values to anchor the music to the user's viewpoint.

3 Implementation

3.1 Octagonal Loudspeaker Layout

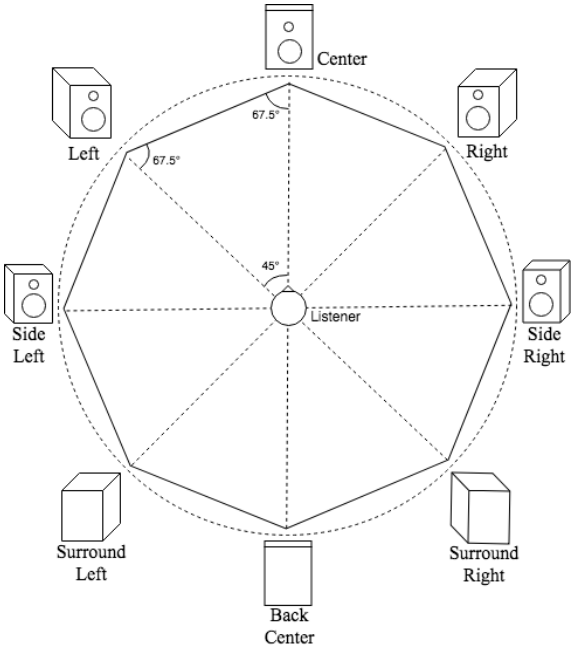

Figure 2. Octagonal loudspeaker layout

An evenly-spaced octagonal loudspeaker layout (Center, Left, Right, Surround Left, Surround Right, Side Left, Side Right, Back Center) in the horizontal layer is chosen, as shown in Figure 2, because of its relatively small separation angle between two loudspeakers (45º), which ensures that the sound source positioning is more accurate compared to a standard 5.1 configuration, for instance, where separation angle between front and rear speakers is ~80-100º. Also, since the user is not stationary, an equally spaced loudspeaker layout is desirable to

ensure a consistent listening experience. The loudspeakers are all positioned at ~1.6m distance from the user.
With respect to the anchored background music, the ideal (reference) direction of the user's gaze is the center channel (0º azimuth). Using a head tracker, the user azimuth can be obtained in real-time, which yields the corresponding azimuth shift required for the background music. Thus, the gain factors of the loudspeakers can be calculated to anchor the music successfully.

3.2 Workflow

The system consists of 3 main components- The Oculus Rift, UnityOSC UDP connection, and the processing in Max/MSP as shown in Figure 3.
The Oculus Rift head tracking data is collected by Unity in real-time, which represents the user's viewpoint in terms of azimuth angle. Using the UnityOSC framework [3], this data is packaged and sent via UDP (User Datagram Protocol) to the Max patch every 1ms. The head tracking data received in Max determines the angle by which the background music needs to be rotated to remain anchored, and is used to calculate the gain values of the speakers using VBAP. Once the speaker gain values are calculated, the 8 channels of music data can be routed to the DAC, providing the anchored music experience.
In this system, the music is upmixed from stereo to a custom 8.1 output setup in real-time using a modified version of the upmixing patch presented in [4]. This output is achieved by modifying the typical 5.1 surround sound configuration output by adding back, side left, and side right channels to obtain an octagonal configuration.

3.3 Crossfade

Depending on the user's viewpoint (azimuth angle), it is required to remap the output channels. When the azimuth angle of the user's view α is between 0º and 45º ($0º < \alpha \leq 45º$), for instance, the phantom center channel is represented through speaker 1 and 2 by VBAP, as shown in Figure 4. When the user's gaze is between 45º and 90º ($45º < \alpha \leq 90º$), the center-

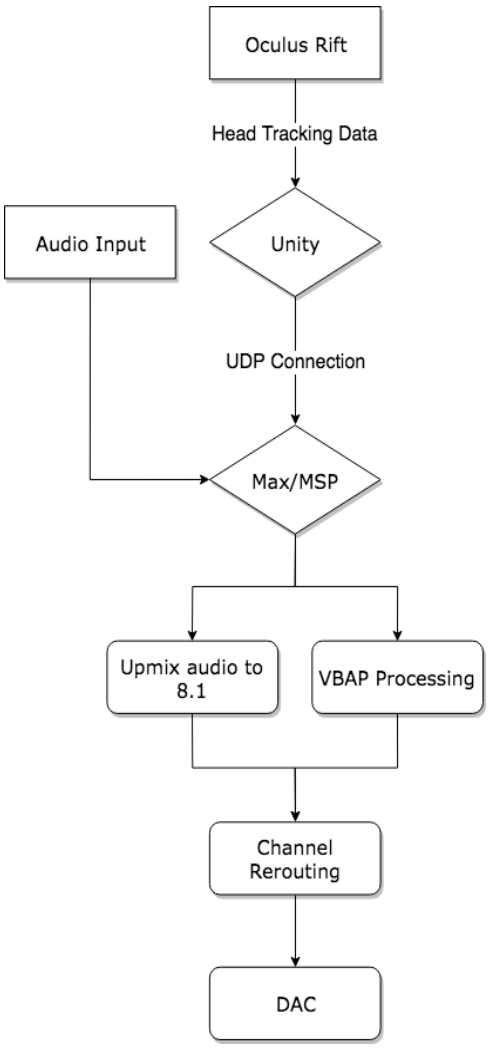

Figure 3. A flowchart of the real-time head tracking process between Unity and Max

channel is represented by speaker 2 and 3. Therefore, when the user's gaze crosses over by 45º, the output channels are rerouted to the corresponding loudspeakers. The rerouting processing could cause a click noise because of instantaneous volume change of the previous signal or immediate attack of the new channel signal. To attenuate these pops, a cross-fader is used when rerouting the output channels.

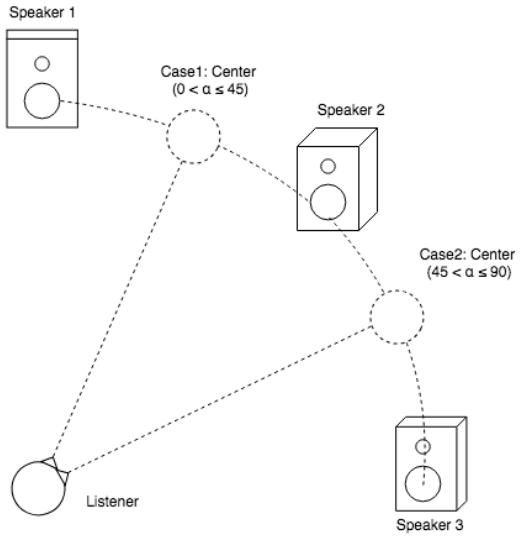

Figure 4. Virtual sources for different user head positions

4 Discussion

The system works well in terms of anchoring the music to the user's viewpoint based on the head tracking data. However, a pop sound was observed when the virtual source shifted from one speaker pair to another. Especially in situations where the virtual source moves quickly between speaker pairs, interpolation might be required to ensure that there is a smooth signal movement from one point to another.

The system proposed here assumes that the speakers are equidistant from the listener, but it is possible to implement a flexible system where speaker channels that are closer to the listener are delayed, compensating for the different propagation times of the speakers [5]. For a flexible speaker layout with an active user, DBAP can also be used. DBAP is essentially an extended version of VBAP where a loudspeaker array of any size can be considered, with the added advantage that the speakers can be in any layout, and their position with respect to the listener and the other speakers is flexible [6]. DBAP is also similar to VBAP in terms of source localization accuracy [7], making it a potential alternative for such setups.

It remains to be seen how the proposed system would affect user experience, particularly in VR environments with a combination of anchored (background music) sources and physically stationary sources (objects in the scene) as well.

5 Conclusion & Future Work

This paper presents a method that uses VBAP to anchor the background music to users' viewpoint in VR applications. The head tracking data is transferred from Unity to Max with the help of a UnityOSC UDP connection. 8.1 channels are established by upmixing the input audio file. The azimuth data obtained from Unity is used to calculate gain factors in real-time, and depending on the user's viewpoint, the output channels are rerouted to corresponding speakers in the 8.1 loudspeaker configuration. Future work might involve adding height layer speakers for a more immersive experience, and conducting a subjective test to study how this system could augment the user experience.

References

[1] Pulkki, V. "Virtual sound source positioning using vector base amplitude panning," *Journal of the Audio Engineering Society, 45(6),* pp. 456-466, 1997.

[2] Roginska, A., & Geluso, P. (Eds.) *Immersive Sound: The Art and Science of Binaural and Multi-channel Audio*. Taylor & Francis, pp. 251-252, 2017.

[3] Martin, J. G. UnityOSC: Open Sound Control Classes and API Interface for Unity 3D (Version 1.2) [Software]. Available from https://github.com/jorgegarcia/UnityOSC, 2012.

[4] Kim, S. "Subjective Evaluation of Stereo-9.1 Upmixing Algorithms Using Perceptual Band Allocation," *Audio Engineering Society Convention 145*, pp. 1-10, 2018.

[5] Pulkki, V. "Generic panning tools for MAX/MSP," *International Computer Music Conference*, 2000.

[6] Lossius, T., Baltazar, P., & de la Hogue, T. "DBAP–distance-based amplitude panning", *International Computer Music Conference*, 2009.

[7] Kostadinov, D., Reiss, J. D., & Mladenov, V. "Evaluation of distance based amplitude panning for spatial audio," *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 285-288, 2010.