Moving Sound Sources over Headphones Via Vector Base Amplitude Panning

Charles Craig Jr. and Tom Xi

3D Audio
Dr. Agnieszka Roginska
May 6th, 2016

Abstract

While Head Related Transfer Functions (HRTFs) offer adequate characterizations of soundfields for mapping virtual point sources to static positions in virtual spaces, they do not offer a comparable mapping for moving sound sources. This paper discusses how through the use of a modified version of vector base amplitude panning, a signal successfully travelled along a user defined path of head related transfer functions. An overview of the methods and procedures used to successfully implement this modified VBAP method are also discussed, as well as areas for improvement.

Background

A goal of the audio reproduction field is to simulate three-dimensional sound fields within a virtual space [2]. Interaural time differences (ITDs) and interaural level differences (ILDs) are dependent upon the position of the listener's head relative to the sound source. Head related transfer functions (HRTFs) are listener dependent spectral filters. When convolved with dry sound sources, HRTFs allow them to assume spatially relevant positions within virtual spaces. Because monophonic and stereophonic reproduction do not offer the ITD, ILD, and spectral cues required to represent three dimensional sound fields, new methods have been proposed, analyzed, and tested. Although each succeeds in enlarging the sound field, pantophonic, holophonic, periphonic, and ambisonic sound positioning systems remain inhibited by fixed loudspeaker positions. [3, 4, 5] Thus, in order to reach the goal of virtual three-dimensional sound field simulation, a more robust and flexible positioning system was developed.

Vector base amplitude panning (VBAP) is a sound positioning system first Introduced by Pulkki et al. in 1996. [1] It allows for the positioning of virtual sound sources within an arbitrary loudspeaker configuration. With each loudspeaker equidistant from the listener, an unlimited array of randomly placed loudspeakers can be used to produce soundfields. [2] Amplitude panning is used to position virtual sound sources between two or three loudspeakers by calculating the gain factors of the sound signal(s) at each location. The use of vector bases allows for the efficient computation of the panning. Because of the flexibility of the VBAP system, the methods, procedures, and results of this new use of the positioning system will also be reviewed.

Motivation

With their $2 billion acquisition of Oculus in 2014, Facebook exhibited an unprecedented degree of faith and foresight in the future demand for immersive virtual experiences by consumers. [6] In 2016, quality consumer priced virtual reality (VR) technologies have begun to saturate the tech market. With forthcoming/available headsets from big name companies such as Sony (PlayStation VR), HTC/Valve (HTC Vive), Samsung (Gear VR), Google (Google Cardboard), and Microsoft (HoloLens), the solidification of VR technologies as viable and profitable entertainment mediums is closer than ever before. [7] Because of this, engineers in the audio reproduction field are striving to improve and apply various sound techniques (introduced by previous generations of audio researchers and engineers) to current VR technologies.

The authors were motivated to position sound sources between measured HRTF points using VBAP because HRTFs had not yet been used to map sound field characteristics to sound sources in motion. Creating a method for sound interpolation between HRTFs, showing distance between the listener and the source, and dealing with virtual point sources in the very-near-field region versus inside one's head were the underlying issues fueling the experiment.

The goal of this project was to create a system which would allow listeners to map a virtual point source onto an arbitrary trajectory in an anechoic space. We aimed to build a demo with predefined trajectories, as well as create a simple interface for listeners to map their own trajectories. The system would process the sound off-line and then output a processed file to be presented to the listeners over headphones.

Methods and Procedures

As has been previously stated, the method chosen to move sound sources over headphones was three-dimensional VBAP. An HRTF data set (downloaded from the IRCAM website), MATLAB scripting, and a 15'' MacBook Pro were used in order to realize the goal.

The procedure for realizing the goal began with the creation of a .txt file from which the MATLAB script could recognize the HRTF files. When run, the script reads in the audio signal and makes note of the user defined path of travel, storing its travel coordinates as points within a vector. The path itself is nothing more than a discrete version of the intended trajectory represented by a sequence of points in space (locations). Points along that path are then used to ascertain the location of the HRTFs closest to the signal as it reaches a point. As the signal arrives at a point, the unit vectors of the three closest HRTFs are found. The signal undergoes windowing at the point location, with each window convolving with the HRTFs closest to that location before being overlapped and added back together. A linear combination of the resulting three basis vectors is calculated to form a vector which points to the current location of the signal. At this point, matrix division must be used to solve for the weights of this linear combination, resulting in a vector which will contain the gain factors of the HRTFs. Normalization of the gains is performed to satisfy the equal power rule and ensure that the resulting signal is correctly represented. Finally, the output signal was written to a file and stored for comparison with the original file.

Results

Using VBAP to interpolate an audio signal along a path of HRTFs proved successful. Signals traveled along the azimuth and elevation planes, with more noticeable differences in position occurring along the horizontal axis than the vertical axis. When listened to by a group of classmates on the day of the presentation, many remarked how they were able to hear the motion when comparing the original audio to the processed audio. When testing the MATLAB script, the authors noticed noticeable coloration and volume differences as the sound signal traveled from point to point. Normalization was employed, however, the output still remained quieter in-between HRTF locations than when they occurred closer to them. Because VBAP is based on the assumption that the user is will be using their head to follow the sound source, we assumed that the normalization issues resulted from this assumption. In order to alleviate them, RMS matching was used in lieu of traditional normalization. As for the variation in spectral coloration, the authors found that the effects were significantly diminished when using an audio file with content other than noise (such as a piano, vocalist, etc.). The lack of discernable differences between the perceived signal position along the elevation plane is most likely due to the small number of points available for the signal to move along (-45° to 45°).

Conclusion

A new method of using VBAP to move sound sources over headphones was introduced and demonstrated. The introduction of HRTFs as loudspeaker positons has been proven to be an effective use of the positioning system. As discussed in the results section, the author's implementation of VBAP was rough around the edges and is in need of tweaking. Future work on the project would include improving the speed and precision of the algorithm which chooses the HRTFs. The introduction of distance information along the z-plane is an improvement which would allow for the signal to move within the sphere in addition to its surface. One of the ways this could be achieved is through the addition of a model which could affect the wet/dry ratio of the signal. Such an addition is important to the viability of using this modification of the positioning system to enhance the impressiveness of audio in VR among other experiences. Improving the scripts runtime is an especially important topic for future work. Currently, a significant amount of time is required for processing of the signal to occur, making this method inefficient for real time applications. Lastly, the authors would like to create an intuitive user interface through which users can map their own trajectories around a sphere of HRTF locations.

Sources

[1] V. Pulkki, J. Huopaniemi, and T. Huotilainen, "DSP Tool for 8-Channel Audio Mixing," in Proc. Nordic Acoustical Meeting 96 (1996 June), pp. 307- 314.

[2] V. Pulkki, "Virtual Sound Source Positioning Using Vector Base Amplitude Panning*" in J. Audio Eng. Soc. Vol. 45, No.6, (1997 June), pp. 456-466.

[3] G. Steinke, "Surround Sound-The New Phase: An Overview," presented at the 100th Convention of the Audio Engineering Society, J. Audio Eng. Soc. (Abstracts), vol. 44, p. 651 (1996 July/Aug.), preprint 4286.

[4] R. Condamines, "Le relief sonore tetraphonique," Rev. d'Acoustique, pp. 277-283 (1972).

[5] M. A. Gerzon, "Periphony: With-Height Sound Reproduction," J. Audio Eng. Soc., vol. 21, pp. 2-10 (1973 Jan./Feb.).

[6] M. Zuckerberg, (2014), website: https://www.facebook.com/zuck/posts/ 10101319050523971

[7] P. Lamkin, (2016), “The Best VR Headsets: The virtual reality race is on” website: http://www.wareable.com/headgear/the-best-ar-and-vr-headsets