4-18-2022

"Contemporary Audio Software Designs and Applications"

Jacob Wolfe
Western Michigan University, Wolfej17@gmail.com

Decreasing Binaural Ambisonic Localization Error with Spherical Harmonic-Based Filtering

Jacob Wolfe
Western Michigan University, Undergraduate, United States of America,
jacob.w.wolfe@wmich.edu

***Abstract* – In this paper, the author presents results from a localization experiment of virtual sound sources on a 2-dimensional 9 channel second-order circular Ambisonics system for acoustic playback, and an identical virtual Ambisonics system created for headphone playback using head-related impulse responses (HRIRs) taken from the acoustic loudspeaker system with a Neumann KU-100 Dummy Head recording system. To reduce subjective localization error with headphone listening, a set of mathematical equations that can be used to parameterize a shelving filter are outlined. These equations include two methods of parameterization; a static method that calculates a constant gain weight based on a degree of localization error, and a dynamic formula that uses the energy value of the closest spherical harmonic component to the localization error degree and scales the resulting value between a minimum and maximum filter gain. This creates a filter that is fully 'activated' when the corresponding spherical harmonic component is fully 'activated.' Results of this study/system yield a multidimensional sound field that contains subjectively more accurate sound localization.**

I. Introduction

A. Background

Spatialization of sound has been a main point of interest for electroacoustic music composers and experimental academics, and the Ambisonics system has recently found expanded use in emerging technologies. Virtual and Augmented Reality (Englund et al., 2017), Electronic Travel Aids (Spagnol, 2018), and previously, aircraft auditory displays (Begult, 1990) have used Ambisonics due to its 'middle ground' nature which includes reproduction techniques by both recording spatialized sound and synthesizing virtual sound fields.

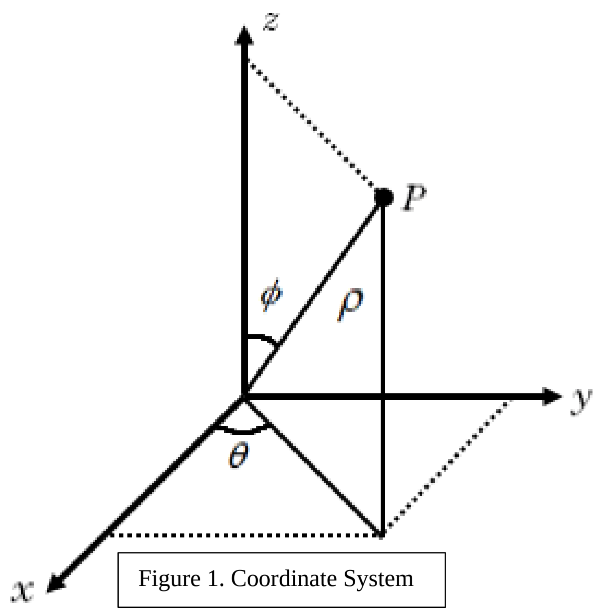


Figure 1. Coordinate System

The basics of Ambisonics can be understood in orders, with the first order consisting of four channels that create a three-dimensional sound field: *0*, *1*, *2*, and *3*. Using spherical polar coordinates as shown in Fig. 1 -, one can encode a signal $s(t)$ at direction (θ, ϕ), with simple ACN ordering equations (F. Zotter, & M. Frank, 2019):

$$0 = s(t)$$
$$1 = s(t)\sin(\theta)\cos(\phi)$$
$$2 = s(t)\sin(\phi)$$
$$3 = s(t)\cos(\theta)\cos(\phi)$$

The sound field can then be decoded for a variety of loudspeaker configurations, comprehensively described by F. Zotter, & M. Frank, (2019). Extending this system to higher orders can be done as also described by Zotter, & M. Frank, (2019), which adds more encoding/decoding signal components. A basic second order systems would include 5 more signal components, 9 in total, and a sound source can be encoded with the following equations:

$$4 = s(t)\sin(2\theta)\cos(\phi)\cos(\phi)$$
$$5 = s(t)\sin(\theta)\sin(2\phi)$$
$$6 = s(t)\,(3\sin(\phi)\sin(\phi) - 1)$$
$$7 = s(t)\cos(\theta)\sin(2\phi)$$
$$8 = s(t)\cos(2\theta)\cos(\phi)\cos(\phi)$$

A simple sampling decoder D can be achieved for the number of loudspeakers L by sampling the spherical harmonics $y_N(\theta)$ at loudspeaker directions $\{\theta_L\}$, and using the matrix $Y_N = [y_N(\theta_L) \ldots]$ in the decoding equation:

$$D = \sqrt{\frac{s_D - 1}{L}}\, Y_N^T$$

The surface of the unit sphere is denoted as $S_2 = 4\pi$, or $S_1 = 2\pi$ as the circumference of the unit circle and the factor $\sqrt{\frac{s_D-1}{L}}$ shows that each loudspeaker synthesizes a fraction of the energy measured on the circle or sphere. 2D and 3D rotation of the sound field around the X, Y, Z axis can be achieved by applying a rotation matrix to the encoded spherical harmonics as described in (Pinchon and Philip), and F. Zotter, & M. Frank, (2019). The sampling decoder exhibits optimal division of energy when the loudspeaker layout is that of a platonic solid with vertices $\mathrm{V} \geq L + 1$. For headphone playback such as that used in Virtual Reality, early works by Jot et al., (1999) outlined the use of impulse responses where each loudspeaker in the Y_N matrix is convolved with the head-related impulse responses (HRIRs) of the corresponding playback directions.

B. Motivation

Noisternig, (2003) states that the use of a generic head for model measurements to encode multi-dimensional sound yields degradation in the localization accuracy for humans. Individuals possess distinct physiological characteristics that determine their perception of sound, making everyone's hearing mechanism a one-of-a-kind system. The dummy head model is used as an "average" to accommodate for many different listeners, each with their own unique head shape. Varying geometries of the body also contribute to the differences in human hearing systems, as described by Rothbucher, (2013). This raises the question; how can the dummy head system be improved with subjective alterations to the encoding Ambisonics system so the improvements may be produced on any virtual loudspeaker configuration? Studying the accuracy of the system for multiple subjects could show how improvements can be applied to create a system that caters more closely to an individual's listening system.

II. Methods

A. Test Environment

For a test environment, the Electronic Music Lab at Western Michigan University was chosen. The measurement system consisted of a 9-object loudspeaker array with 40° of spacing between each speaker and each speaker 7 feet from the central listening point. The Ambisonics system in use is 2D second order encoding with SN3D normalization, ACN channel ordering, and a basic sampling decoder (Zotter, 2019).

B. Stimulus

Out of the many localization stimuli available, broadband pink noise was selected. For each test location, 3 short

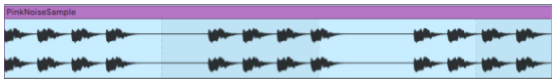


Figure 2. Pink noise test for one location

bursts of the noise spaced by 218ms, with each burst containing 4 transients spaced by 108ms from the previous transient were played.

The cadence of the stimulus is based on previous localization experiments carried out by Pulkki, (2001).

C. Listeners

3 subjects participated in this experiment. The population included all males, aged 20 to 25.

D. Procedure

The perceived direction of sound is measured by pointing the swiveling arm of a protractor, as the overall error of aiming at visual objects has been found to be less than 0.5° according to Blauert's work, as cited in Zotter, (2008). Each stimulus trigger X was played at 30° increments around the azimuthal plane in the range $0 \leq X < 360,$ twice, totaling 24 locations per listening system in a round (48 stimuli total). Each subject did a test round to get familiar with the procedure.

E. Task

The remainder of this paper introduces the improvements, outlines the initial test experiment results with no improvements, and compares initial results with the improvement equations applied.

III. Results

A. Loudspeaker vs Headphones

The perceived direction on headphones does not correlate with complete accuracy to the perceived direction on loudspeakers, and both mediums exhibit localization error based on the reference value.

Localization error on headphones is especially apparent at $0°\theta, 30°\theta, 300°\theta$ and lateral regions.

B. Initial Improvements

Initial attempts included amplitude modifications to the nearest Y_N^T spherical harmonic component at localization degrees of interest (varied by subject). Amplitude of the Y_N^T component was increased for localization degrees of error between $0°\theta$ - $90°\theta$ and between $270°\theta$ - $360°\theta$, and decreased for degrees between $90°\theta$ - $270°\theta$. After one round of Y_N^T amplitude modifications, it was concluded that localization was not improved, and a new improvement method would need to be conceived.

C. Spherical Harmonic-Based Filtering

Early works by Batteau, (1967) describe how the pinna works as a filter to allow discernment of frontal and posterior sounds with equal IID (interaural intensity difference) and ITD (interaural time difference). By increasing high frequency content in frontal sounds and attenuating high frequency content in lateral sounds, one can assume a remedy for incorrect discernment of frontal and lateral sounds with equal IID and ITD. Spherical harmonic components Y_N^T in the first order can be parsed into positive and negative energy and then formatted into a set of gain weights for a filter based on direction using the following equation for positive energy:

$$GN = abs(Y_N^T) + Y_N^T$$

And for negative energy:

$$GN = abs(Y_N^T) - Y_N^T$$

Y_1^1 in this case would have maximum positive energy at $0°\theta$ and maximum negative energy at $180°\theta$. The result of both should then be scaled between a minimum and maximum $G(f)$ gain weight value for a suitable shelving filter, such as the second order Butterworth shelving filter with Q value .707 and cutoff at 3000hz used in this experiment. If a dynamic gain-weight system is not preferred, a static one may be used. The Y_N^T spherical harmonic component at localization degrees of interest should then be filtered at the calculated or set gain value.

D. Proof

Pictured to the left is an average graph of all subject's localization data taken on loudspeakers and headphones without the improvements applied. Pictured to the right is an identical loudspeaker and headphone graph portraying localization direction with the improvement equations applied. The position of the dashed lines verses the solid lines indicates variance between perceived headphone and loudspeaker sound locations, with improvements indicated by closer correlation of dashed to solid lines with the same color.

Figure 3. Headphone (dashed lines) and Loudspeaker (solid lines) no improvements

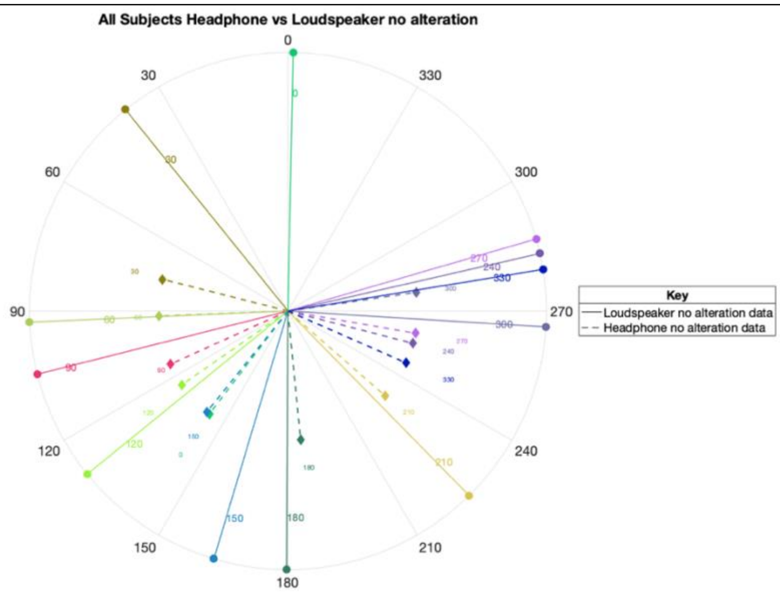
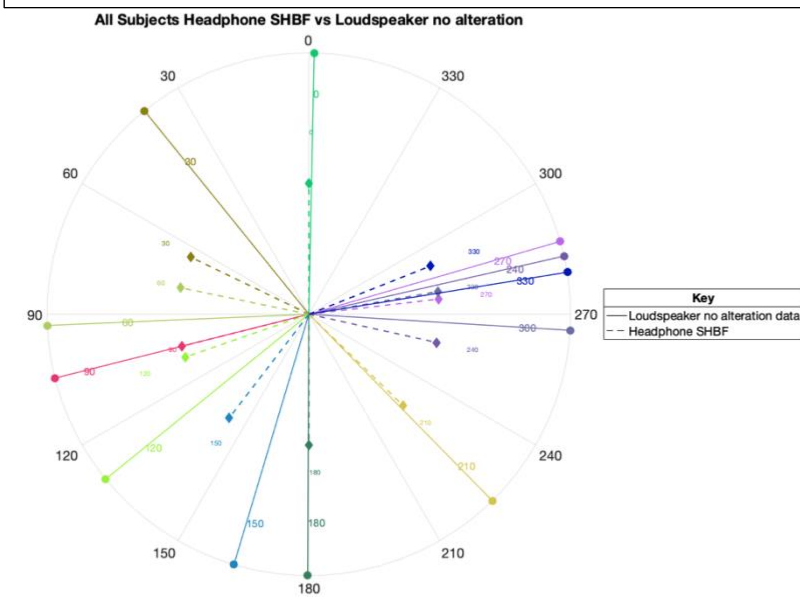


Figure 4. Headphone (dashed lines) and Loudspeaker (solid lines) with SHBF (Spherical Harmonic-Based Filtering)



IV. Future Work

A. Higher Order Ambisonics

This system could be adapted to higher order Ambisonics using static gain weights, and a higher order system would lead to a more spatially focused filtering system (smaller area of filtering on the virtual sphere).

B. Integrated Test

To discern problematic sound localization areas on a binaural Ambisonic headphone system, a pre-packaged test like this experiment could be included with the headphones. Depending on the Ambisonic order in use, static or dynamic gain-weighting of the filter would need to be used.

V. Conclusion

A. Discussion

An ongoing and contemporary spatial audio topic, the discovered improvements address the need for improved binaural localization of sound without the solution being tied to a certain speaker configuration or Ambisonic order.

B. Acknowledgement

The author would like to thank Dr. Christopher Biggs for mentorship and the

Lee Honors College at Western Michigan
University for providing funding.

References

C. Englund, A. D. Olofsson, L. Price, (2017). Teaching with technology in higher education: understanding conceptual change and development in practice, *Higher Education Research & Development*, 36(1). doi: 10.1080/07294360.2016.1171300

Batteau D. W. (1967). The role of the pinna in human localization. *Proceedings of the Royal Society of London. Series B, Biological sciences, 168(1011), 158–180.* doi: https://doi.org/10.1098/rspb.1967.0058

J. -M. Jot, V. Larcher, & J. -M. Pernaux, (1999). A comparative study of 3-d audio encoding and rendering techniques, in *AES 16th International Conference.* Retrieved from http://www.aes.org/e-lib/browse.cfm?elib=8029

M. Noisternig, T. Musil, A. Sontacchi, & R. Höldrich, (2003). 3D binaural sound reproduction using a virtual ambisonic approach. 174 - 178. doi: 10.1109/VECIMS.2003.1227050.

M. Rothbucher, et al., (2013). Comparison of head-related impulse response measurement approaches. *The Journal of the Acoustical Society of America 134*(2). doi: https://doi.org/10.1121/1.4813592

S. Spagnol, et al., (2018). Current Use and Future Perspectives of Spatial Audio Technologies in Electronic Travel Aids. *Wireless and Mobile Computing.* doi: https://doi.org/10.1155/2018/3918284

Pulkki, Ville. (2001). Localization of Amplitude-Panned Virtual Sources II: Two- and Three-Dimensional Panning. *J. Audio Eng. Soc. 49. 753–767*. Retrieved from https://www.researchgate.net/publication/237067976_Localization_of_Amplitude-Panned_Virtual_Sources_II_Two-_and_Three-Dimensional_Panning

F. Zotter, & M. Frank, (2019). *Ambisonics: A practical 3d audio theory for recording, studio production, sound reinforcement, and virtual reality.* Springer.

F. Zotter, M. Frank, (2008). Localization Experiments Using Different 2D Ambisonics Decoders. Paper presented at the 25th Tonmeistertagung – VDT International Convention, November 2008. Retrieved from https://www.researchgate.net/publication/262825665_Localization_Experiments_Using_Differe nt_2D_Ambisonics_Decoders/citation/download