Distance Encoding in Ambisonics Using Three Angular Coordinates

Rui Penha
INET-MD / University of Aveiro, Portugal, ruipenha@ua.pt

***Abstract* — In this paper, the author describes a system for encoding distance in an Ambisonics soundfield. This system allows the postponing of the application of cues for the perception of distance to the decoding stage, where they can be adapted to the characteristics of a specific space and sound system. Additionally, this system can be used creatively, opening some new paths for the use of space as a compositional factor.**

I. Introduction

Sound spatialization has been one of the main aspects of interest for composers since the beginning of electroacoustic music experiments. Several techniques for spatialization were created specifically for synthesizing virtual soundfields, whilst others were derived from sound recording techniques. Amongst the latter, the Ambisonic system [1] has resurfaced in the last years, due to an increasing interest in its characteristics [2] and its advantages over other surround systems [3].

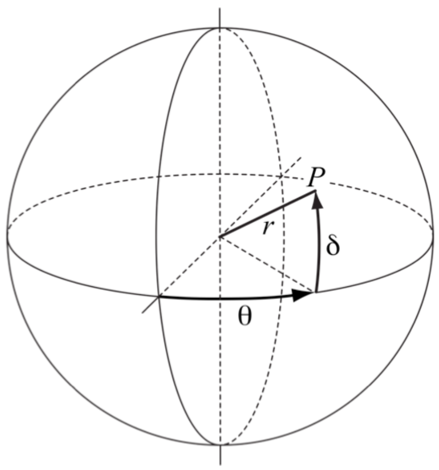

Fig. 1. Coordinate system used in this paper.

The basic first order Ambisonic system is known as the B-format, in which a full three-dimensional soundfield is decomposed in spherical harmonics and encoded into four channels, known as *W*, *X*, *Y* and *Z*. Using spherical polar coordinates - azimuth θ and elevation δ, as shown in Fig. 1 -, one can encode a signal *i* at point *P*, in a first order Ambisonics B-format, using simple equations:

$$W = i\frac{1}{\sqrt{2}},$$

$$X = i\cos\theta\cos\delta,$$

$$Y = i\sin\theta\cos\delta,$$

$$Z = i\sin\delta,$$

using the standard weighting of the 0th order *W* channel [4]. This soundfield can then be reconstructed with a variety of speaker arrays [5], including, with additional processing, binaural stereo [6], using the decoding equation for the speaker signal *s*:

$$s = \frac{1}{S}\left(\frac{W}{\sqrt{2}} + X\cos\theta\cos\delta + Y\sin\theta\cos\delta + Z\sin\delta\right),$$

where *S* represents the total number of speakers and both the azimuth θ and elevation δ represent the angles of the speaker position on the surface of the sphere defined by the concentric speaker array.

Several higher order Ambisonic systems have been developed since the original proposal of the system [4][7], augmenting the spatial resolution of the Ambisonics encoding by increasing the order of the spherical harmonics used to encode the soundfield.

II. Distance Encoding

A. State of the Art

By relying solely on two angular coordinates - azimuth θ and elevation δ - for the encoding, the traditional Ambisonic system only encodes a two-dimensional spherical coordinate system, represented by the surface of the sphere were the sounds are projected. This system, created initially for recording rather than the encoding of synthesized sources, therefore favors the encoding of the localization of sound, excluding the distance cues that are not present or permanently added to the signal at the time of encoding, as it has been proposed with the inclusion of near field compensation at the encoding stage [8].

Additionally, a sound at the exact center of the coordinate system, hence with no azimuth nor elevation and consequently absent from all the spherical coordinates except for the 0th order *W*, cannot be easily encoded with a traditional Ambisonics approach, as addressed with the creation of W-Panning [9] for the spatialization of sounds enclosing the listener.

For higher order Ambisonics, a new encoding system [8] effectively addresses the loudspeaker near field effect by compensating for it at the encoding stage, thus allowing the reproduction of sources both inside and outside the surface of the sphere defined by the speaker array. Furthermore, if the diameter of a given speaker array is different from the reference one, compensating filters can be applied prior to the decoding stage, thus enabling the diffusion of the same encoded soundfield using different speaker setups.

B. Proposed Solution

The proposed solution to encode distance in an Ambisonic-based system consists in encoding the distance r as a new angular coordinate, using the hyperspherical coordinates of a 3-sphere [10], effectively turning the radial coordinate r in the angular coordinate ρ. By varying the angle ρ between 0 (at the center of the sphere) and $\pi/2$ (at the surface of the sphere) and adding one audio channel D, one can create an extended B-format with distance encoding for signal i at point P using the equations:

$$W = i\frac{1}{\sqrt{2}},$$

$$X = i\cos\theta\cos\delta\sin\rho,$$

$$Y = i\sin\theta\cos\delta\sin\rho,$$

$$Z = i\sin\delta\sin\rho,$$

$$D = i\cos\rho.$$

This soundfield with distance encoding can then be decoded using the equation:

$$s = \frac{1}{S}\left(\frac{W}{\sqrt{2}} + X\cos\theta\cos\delta\sin\rho + Y\sin\theta\cos\delta\sin\rho + Z\sin\delta\sin\rho + D\cos\rho\right).$$

As a result, if a signal is encoded with the maximum distance, thus at the surface of the sphere assumed to enclose the soundfield, the D channel is silent and all the others work as in their original form. Conversely, if a signal is encoded with distance $\rho = 0$, thus at the origin of the coordinate system, the D channel receives the signal with its full amplitude and all the other channels are silent, therefore loosing all the localization cues, as the signal is at the listener's position. All the versatility of traditional Ambisonics is retained, as the W, X, Y and Z channels are exactly the same as they would be on a regular system, as long as all the signals are encoded with distance $\rho = \pi/2$. As the D channel only encodes distance and not direction, the traditional matrices for rotation (around the z-axis), tilt (around the x-axis) and tumble (around the y-axis) [4] can still be used with the X, Y and Z channels alone.

Although the diameter of the sphere assumed at the encoding stage must be known when reconstructing the soundfield, its value can now be different than the one of the final speaker array. It is important to note, however, that one needs to decode the central position to either a real or a virtual omnidirectional speaker, the latter being then spread by the real speakers in a weighted manner. Failing to do so would cause a progressive loss of the signals encoded towards the center of the coordinate system, in a similar way as one looses the signals encoded with an elevation angle δ of $\pm\pi/2$ in a horizontal-only speaker array.

C. Distance Cues

Besides allowing the playback of strict distances between sounds in differently sized sound systems, the encoding of the distance between virtual sound sources and the center of the coordinate system can postpone the application of cues for the perception of distance - such as the loudness, atmospheric absorption and reverberation - to the decoding stage, as long as one knows the sphere size assumed during the encoding of the soundfield. This opens the door to the fine-tuning of these cues to each sound projection space from the same encoded signals.

A simple method for decoding, e.g., an extended B-format with distance encoding for a speaker array with a smaller diameter than the one of the assumed sphere would be: to decode the signals for a virtual speaker array, with the same number of speakers as the real one but with the same diameter as the assumed sphere; to decode the signal for a virtual omnidirectional speaker at the center of the assumed sphere; to apply the required distance cues to all the decoded signals, compensating for the difference in diameters between the virtual and the real speaker arrays; to play the compensated signal of each virtual speaker in the concentric array using the real speaker with the same angular location and the weighted signal of the virtual center speaker through all the real speakers.

The nature of the angular coordinate system encoding by itself caters for the smooth fades between the loudness and atmospheric absorption filters applied to the virtual center signal and to the virtual speakers located on the surface of the assumed sphere. Preliminary listening tests have shown that this robust system is capable of very convincing results, if not physically accurate ones, which, although unlikely, remains to be tested.

III. Spatialization Vocabulary

Regardless of the fact that Ambisonics excels at physically reconstructing a recorded soundfield [2][3][7][8], both traditional and new vocabulary for spatialization can be implemented using its standard techniques. The rotation, tilt and tumble matrices, used to alter the microphone position within a recorded B-Format, can be used to continuously rotate a soundfield for creative spatialization proposes. As an example, if one implements some kind of time-dependent processing with a feedback loop, as a reverb or a delay, to an encoded soundfield and rotates the resulting soundfield while rotating the original in the opposite direction, a moving tail is created. A single parameter - the angular rotation velocity, responsible for both the direction and the size of the moving tail - can be used to manipulate the resulting effect in real-time.

By inserting stages of decoding and re-encoding around effects affecting only specific spaces, one can create small spaces within the soundfield where sounds are transformed solely by "traveling" through them. An implementation of this (in higher order Ambisonics) using Max/MSP is shown in Fig. 2. Again, the nature of the angular coordinate system encoding by itself caters for the smooth fades between different areas.

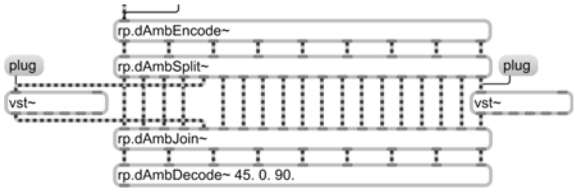

Fig. 2. Applying effects to small areas of the encoded soundfield**.**

The “rp.dAmbJoin~” object shown in Fig. 2 is in fact nothing more than a group of 19 encoders for specific positions within the assumed sphere: elevations $\delta = \pi/2$ and $\delta = -\pi/2$, 16 equally spaced azimuths, all with distance $\rho = \pi/2$, and the center with distance $\rho = 0$. This object can therefore be used to directly encode signals from fragmented sources, e.g. spectral spatialization [12] or streams of grains in granular synthesis.

The above approaches, albeit profiting from the distance encoding, are nevertheless possible to implement using the existing Ambisonic system. Some specific creative approaches, on the other hand, arise from the described proposal. As an example, the diameter of the assumed sphere is employed as a constant while decoding a soundfield for different speaker arrays. If used as variable, however, one can manipulate the perceived size of the diffusion space and the relative distance of all the sound objects encoded in a soundfield in real-time, e.g. by varying the size of a rotating soundfield proportionally to its angular velocity. This effect, while certainly possible to implement using existing techniques, would involve the simultaneous manipulation of numerous effects and parameters, as opposed to solely two variables, when using the proposed solution.

IV. Future Work

A. Higher Order Ambisonics

This system can naturally be expanded to higher order Ambisonics, therefore allowing greater spatial resolution. The proposed standard for the system being developed uses a stream of nine audio channels with a mixed-order system of 3rd horizontal order and 1st vertical order, thus reflecting the ubiquity of horizontal-only speaker arrays while preserving the full-sphere capabilities, with the lower resolution justified by the poorer precision of localization in the median plane [11].

B. Compatibility

The viability of higher order Ambisonics with distance encoding is conditioned by the research of a way to compensate for the near field effect of speakers [8] at the decoding stage.

The relative size of sound sources is also a very important parameter when working with distance encoding. As in the vast majority of sound spatialization techniques, the proposed system encodes sounds as dimensionless points in space. This will be addressed by researching ways to integrate the O-format [9], an Ambisonics B-format that encodes a pattern of sound radiation of an object, within the proposed system.

C. Develpment of Modular Tools for Spatialization

Modular tools for the field application of the proposed system are being created, both as Max/MSP abstractions and externals and as standalone applications. The modular approach for the creation of these tools will allow the scalability of systems within a consistent approach. At the same time, some examples of the specific vocabulary for interacting with the system are being composed. While trying to integrate some of the current Ambisonics-related research into the proposed system, the focus on the creation of straightforward, yet powerful, means of interacting with the tools under development will most certainly require a proposal for a new, modular environment for electroacoustic composition where spatialization plays the chief role.

V. Conclusion

An ongoing research, the proposed Ambisonics system effectively addresses the need to integrate distance cues in the spatialization of sound without being tied to a specific speaker array.

Acknowledgment

The author would like to thank his supervisor, João Pedro Oliveira, for his commitment to this project and inspirational work as a composer, as well as the University of Aveiro for supporting his work.

References

[1] M. Gerzon, “Periphony: with height sound reproduction,” *Journal of the Audio Engineering Society*, vol 21:1, pp. 2-10, January/February 1973.

[2] J. Bamford, *An Analysis of Ambisonic Sound Systems of First and Second Order*, unpublished MSc thesis, 1996.

[3] D. Malham, “Homogeneous and nonhomogeneous surround sound systems,” *AES UK “Second Century of Audio” Conference*, London, UK, June 1999.

[4] D. Malham, *Higher order Ambisonic systems*, http://www.york.ac.uk/inst/mustech/3d_audio/higher_order_ambisonics.pdf, retrieved in April 2008.

[5] F. Hollerweger, *Periphonic Sound Spatialization in Multi-User Virtual Environments*, Santa Barbara: CREATE, 2006.

[6] M. Noisternig, A. Sontacchi, T. Musil and R. Höldrich, “A 3D ambisonic based binaural sound reproduction system,” *AES 24th International Conference*, Banff, Canada, June 2003.

[7] J. Daniel and S. Moreau, “Further study of sound field coding with higher order ambisonics,” *AES 116th Convention*, Berlin, Germany, May 2004.

[8] J. Daniel, “Spatial sound encoding including near field effect: introducing distance coding filters and a viable, new ambisonic format,” *AES 23rd International Conference*, Copenhagen, Denmark, May 2003.

[9] D. Menzies, “W-panning and O-format, tools for object spatialization,” *AES 22nd International Conference of Virtual, Synthetic and Entertainment Audio*, Espoo, Finland, June 2002.

[10] Wikipedia, “3-Sphere,” http://en.wikipedia.org/wiki/3-sphere, retrieved in April 2008.

[11] G. Kendall, “A 3-D sound primer: directional hearing and stereo reproduction,” *Computer Music Journal*, vol 19:4, pp. 23-46, Winter 1995.

[12] R. Torchia and C. Lippe, “Techniques for Multi-Channel Real-Time Spatial Distribution Using Frequency-Domain Processing,” *Proceedings of the 2004 Conference on New Interfaces for Musical Expression*, Hamamatso, Japan, June 2004.