Audio Engineering Society
Conference Paper
Presented at the Conference on
Spatial Reproduction
2018 August 6 – 9, Tokyo, Japan

Controlling The Apparent Source Size In Ambisonics Using Decorrelation Filters

Timothy Schmele[1,2] and Umut Sayın[1]

[1] *Eurecat, Centre Tecnològic de Catalunya, Multimedia Technologies Group, Barcelona, Spain*
[2] *University of Music Karlsruhe, Institute for Musicology and Music Informatics, Karlsruhe, Germany*

Correspondence should be addressed to Timothy Schmele (`tim.schmele@ce.eurecat.org`)

ABSTRACT

The apparent source extent, defined as the perceived sound source size, is a common parameter for 3D audio applications, such as VR, 3D cinema and spatialized music composition. Here, we present a new method for achieving this effect for Ambisonics, which targets the minimization of the interaural cross-correlation coefficient. This is achieved by reducing the Ambisonic order and rotating decorrelated copies of the original stream into an equidistant configuration. Comparisons to previous methods as well as different filter parameters and parameter curves are discussed for both binaural and multichannel speaker systems. Preliminary results indicate that this method is able to stable images, particularly in the gradual change in source size and listening area stability.

1 Introduction

In this paper, we describe a new method for controlling the apparent source extent or perceived sound source size in 3D auditory displays. Naturally, not all sound sources in real life are point-like but exhibit a greater auditory size [1, 2]. Hence, such a control parameter is a desirable feature in 3D audio applications, including VR, cinema or musical works [3, 4].

The effect can be attributed to different spatial features, mainly the inter-aural cross-correlation coefficient (IACC) and non-spatial acoustic features such as the loudness, bass content and the duration of a source [2, 5]. Here, we propose a new method that modifies the IACC for both multichannel loudspeaker arrays and binaural rendering and which is particularly suited to be implemented in the Ambisonics domain.

The most straightforward method to distribute a sound is to play the source sound through as many loudspeaker or binaural points as necessary [6]. While this method can work in binaural displays where the listener remains in the sweet-spot, a listener in a multichannel loudspeaker array may be closer to one speaker, causing the sound to collapse into it.

Some methods try to overcome this problem by either distributing the sound in frequency or time. Methods that place different frequency bins into different directions have been shown to perform poorly when tested perceptually [1, 3], although efforts have been made to improve this approach [7]. Granular techniques that exploit a distribution in time may work well with percussive sounds like rain, but could exhibit artifacts with sustained sounds [3].

Other methods that target the minimization of the IACC directly introduce phase shifts on frequential sub-bands [2, 3, 8, 9]. Generally it is proposed to use one filter per loudspeaker, limiting the solution to the specific loudspeaker layout it is rendered on.

Specifically considering Ambisonics, two principal methods have been proposed. The first blurs any Ambisonic signal by rotating it within the range of the angular blur parameter [10]. This method can be considered similar to granular distribution techniques, as a distribution by fast rotation.

A second method tries to take advantage of the regular blur that occurs at lower orders in Ambisonics [4]. Maintaining the overall energy, it reduces the Ambisonic order gradually with increased desired source width. Because the change in blur is less in higher orders than in lower ones, a compensation curve needs to be introduced to counteract the skew. Moreover, this method results in a correlated signal over all speakers when reaching the 0^{th} order Ambisonics [4], resulting in the same shortcomings as the straightforward method described above.

2 Methods

We propose a new method which combines several aspects of most of the methods described above. The order reduction method [4] is first modified for SN3D and serves as a basis on which we intend to improve on. This also allows us to concentrate on low orders of Ambisonics, preferably in the 1^{st}, which also saves computational cost.

The principal idea is to distribute decorrelated copies of the signal in an equidistant configuration around the listener to minimize comb filtering. Because 1^{st} order Ambisonics already exhibits a natural blur, we can make use of this and only require a minimum of decorrelated copies in a uniformly distributed configuration to cover the whole sphere.

To achieve a smooth transition from minimum to maximum source size we reduce the Ambisonic order gradually down to the 1^{st} while also gradually increasing the volume of the decorrelated copies as well as rotating them from the original source position to the final uniformly distributed configuration. Nonlinear mapping can help to adjust the transitions between methods and aid in a perceptually more robust image.

2.1 Gain Correction in Ambisonic Order Reduction

As a starting point, the Ambisonic order is reduced gradually as a gain factor $g_n(\alpha)$ with $n \in [0,N]$ is applied to the Ambisonic signals where N is the Ambisonic order of the signal, n is the subsequent lower orders for the gains to be applied. The reduction is applied to each order as a function of increasing spatial blur parameter $\alpha \in \mathbb{R} \bigcap [0,100]$,

$$g_n(\alpha) = 1 - \frac{1}{1+e^{-\tau(\alpha - 100\frac{N-n+1}{N+1})}} \tag{1}$$

resulting in a fade out of higher order components [4]. τ is provided as a smoothing factor for the fade-out effect. These gains are later subjected to an energy preservation factor $W(\alpha)$

$$W(\alpha) = \sqrt{\frac{\Sigma_e(0)}{\Sigma_e(\alpha)}} \tag{2}$$

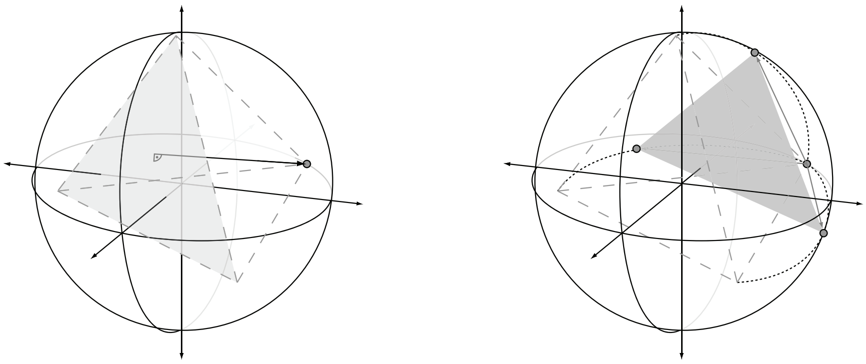
where Σ_e denotes the decoder energy

$$\Sigma_e(\alpha)_{N3D} = \Sigma_{n=0}^{N}(2n+1)g_n^2(\alpha) \tag{3}$$

in a three dimensional loudspeaker setup with N3D normalization. For a SN3D normalization it is necessary to readjust the energy preservation by a factor of $\sqrt{2n+1}$ for each order resulting in a decoder energy function of

$$\Sigma_e(\alpha)_{SN3D} = \Sigma_{n=0}^{N}(\sqrt{2n+1})g_n^2(\alpha) \tag{4}$$

As stated in [4], $\alpha = 100$ should be avoided due to the fully correlated signal problem that is mentioned earlier. Furthermore, spatial blur through sweet spot reduction by Ambisonic order decrease was deemed too weak by itself as an angular source size effect by the authors. Hence, this method was further extended using spatially distributed decorrelated copies of the same sound source.



(a) (b)

Fig. 1: Visualizations of the decorrelated copy movement in a tetrahedron configuration. Figure (a) shows the tetrahedron in dotted lines and the normal vector to the original source which defines the distance of the decorrelated copies. (b) shows the path of movement (curved dotted lines) as well as an intermediary position of the decorrelated sources for some $0 < \alpha < 100$.

2.2 Source Decorrelation Distribution

In order to increase the decorrelation of a spatially extended source, decorrelated copies of the same sound are distributed around the listening space. These copies are gradually introduced as the source extent parameter increases. The introduction happens in two forms: through gain and position.

In gain, the decorrelated copies are gradually faded in with the increase of the source extent α, while the original source amplitude should be decreased to match the loudness between all at full source extent. Because the decorrelated copies are altered in phase randomly respectively (see section 2.3), they can be considered incoherent to each other and to the original source. For $m-1$ decorrelated copies, the final gain should hence be $\sqrt{1/m}$.

In terms of positioning, the following criteria were defined:

- The number of copies should be kept at a minimum to reduce computational effort as much as possible
- The number of copies should be high enough to achieve adequate coverage over the sphere considering the Ambisonic order used
- The copies should be uniformly distributed
- The distance should be far enough to avoid comb filtering, due to the random phase shifts

The method described in section 2.1 would potentially reduce the Ambisonic order down to the 1st. Using the *energy vector magnitude* [11], also known as *Gerzon localization vector*, it follows that 1st order Ambisonics exhibits an angular source size of about 135° to 180° [4, 12]. Considering that any Ambisonic discretization of order N requires a minimum of $(n+1)^2$ speakers to adequately be represented [13], we can assume that in 1st order Ambisonics we would require a minimum of 4 points, i.e. 3 decorrelated sources to adequately distribute a signal across the sphere with minimum overlap.

The regular tetrahedron therefore was chosen as the target shape not only for it's regularity in the Ambisonic sense [13], but also for it having the least vertices of any volume. It is rotated so that the three decorrelated sources form a triangle perpendicular to the original source position (see figure 1(a)).

As the source extent parameter α is increased, the copies are rotated to increase their angular distance from the original source direction until they match the final positions of the rotated tetrahedron. This this can be implemented as an interpolation in rotation. Although figure 1(b) shows the decorrelated copies as points, this method can easily be applied to copies of the whole 1st order Ambisonic representation of a scene, making it feasible as a post-production tool.

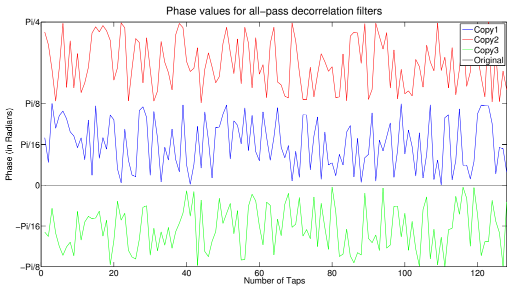


Fig. 2: Example of random phase data of the decorrelation filters for a tetrahedron distribution with K=128 taps. The original signal (black) is kept non modified while the other three copies(red,blue and green) are filtered by random phase all-pass filters with alternating offset and constant range.

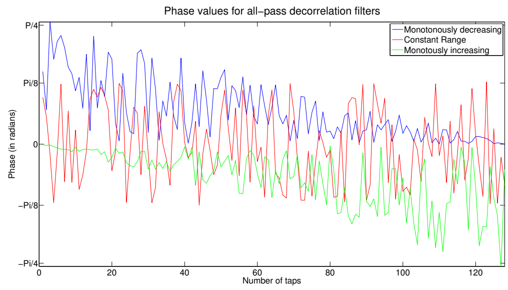


Fig. 3: Example of different type of random phase data of the decorrelation filters with K=128 taps. The filters can be either monotonously increasing (green), constant range (red) or monotonously decreasing (blue)

For comparison, we also chose to implement two distributions with a 2^{nd} order reduction limit. Applying the above considerations, a 2^{nd} order Ambisonic signal would require a minimum of 8 decorrelated copies plus the main source. Because there is no platonic solid with 9 vertices, we constructed both a uniformly distributed layout using the method described in [12] as well as a dodecahedron.

2.3 Decorrelation Filter Construction

The decorrelated copies are generated by filtering the original signal with all-pass FIR filters with random phase. For a filter with $k = 0, 1, 2...K$ taps the transfer function is defined by

$$H_m = A_m(k)(cos(\Phi_m + i(sin(\Phi_m) \quad m = 1, 2, ..M \quad (5)$$

where $\forall k \rightarrow A_m(k) = 1$ (all-pass) and Φ_m is the corresponding random phase vector. Even though the random phase can be generated through any random number generator (RNG), in this case a Mersenne Twister algorithm is deployed [14] to generate the phase vector in form of

$$\Phi_m(k) = (b-a)RNG_{Mersenne}(k) + a; \qquad (6)$$

where coefficients a and b presents a range and and an offset to the randomized phase vector.

The impulse response (IR) $h_m(t)$ is then calculated by taking the inverse Fourier transform of the transfer function H and the source signal is convolved with the resulting IR. Different sets of filters are generated for the listening test where the filters are provided with constant/alternating offset and/or constant/alternating range as demonstrated in Figure 2.

To test the frequency dependency of the source size perception monotonously increasing/decreasing filters are generated where the range of modification of the phase is different in different ranges as demonstrated in Figure 3. Also, filters of different lengths of 128 or 512 taps were constructed and compared.

2.4 Mapping α

When combing the methods discussed above, we are presented with several parameter space dimensions that can be mapped to other inclinations or nonlinearly. First, the angular spread of the gain correction in Ambisonic order reduction method, as discussed in section 2.1, is not linear in angular spread with respect to α, which can be corrected for with a nonlinear mapping function $g_n(\alpha)$ [4].

Furthermore, each dimension in the parameter space can be given preference at different ranges of α. For example, due to the nonlinear angular spread behavior, the effect of the Ambisonic order reduction is relatively weak in higher order. Yet, introducing the decorrelated copies too early and too close to each other exhibits comb filtering, which is to be avoided. In order to defer the introduction of the decorrelated copies, the

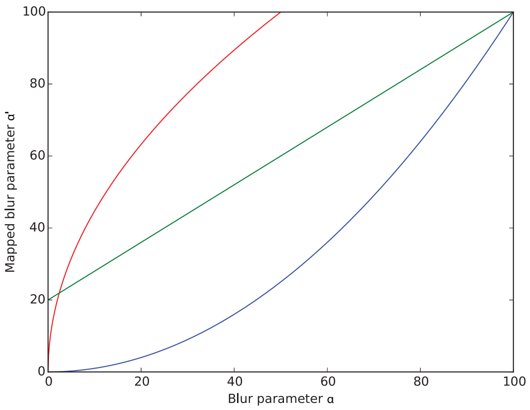


Fig. 4: Example of a parameter space mapping. The Ambisonic order reduction (red) is accelerated for low values of α and inclined to arrive at maximum early ($\alpha' = min(\sqrt{2\alpha}, 100)$). The copied source's position (green) is offset with an adjusted inclination and the copied source's gain (blue) is increased with a slow start to counteract comb filtering.

Ambisonic order reduction method can be accelerated. It's strength in lower orders can then aid in bridging the gap until the deferred gain raise of the decorrelated copies becomes audible at a later stage. Additionally, the position of the decorrelated copies can be set to already start apart by introducing a positive offset to the mapping (see figure 4).

3 Discussion

Preliminary tests were done in both binaural and a multichannel speaker configuration with 25 speakers in a 2/3-sphere. Comparisons were done relating this new method to the simpler Ambisonic order reduction method [4]. Furthermore, the 1st order variant of this new method was compared to the both 2nd order variants with 8 and 11 filters respectively. In each case, we also compared the different filter types in terms of their decorrelation strength as well as tendency to produce comb filter effects.

As mentioned earlier, the order reduction approach is very subtle in higher order ranges. Effects were only really audible when reducing from 2nd through 1st and down to the 0th order. Also, when reducing orders the source is perceived as moving instead of increasing in size in multichannel speaker layouts. This is most likely due to correlation effects of speaker feeds and the listeners relative distance to each speaker, as the signal tends to be fully correlated in all speakers in the case of the 0th order.

According to preliminary tests, the proposed method improves the sensation of gradual change in source size while providing more spatial stability at full source extent when compared to the simple order reduction method. A gradual transition from a point source to a fully distributed source can be achieved, with an impression of constant size increase within intermediary values. Traversing through the space in a multichannel speaker setup when listening to a fully spread source gives an impression of a sound present everywhere and from all sides.

Increasing the number of decorrelated copies improves the stability of the perceived source size in multichannel loudspeaker setups without altering the listening area. But, when compared through headphones, it also seemed to reduce the externalization in binaural rendering. Also, adding more decorrelated copies made it more difficult to avoid comb filter effects, especially at small source sizes with low values α.

We observed that the volume and displacement velocity of the decorrelated copies can help avoid comb filtering effects, which occurs due to the overlapping main lobes of the decoded signals. It is required to separate the copies far enough before fading in, while simultaneously avoiding a discontinuity in spatial extent. Further testing will be done for adjusting these parameters for a real life application where both binaural listening and multichannel loudspeaker setups are considered.

The overall sensation of change in source size is different in cases of binaural and multichannel loudspeaker listening. In a loudspeaker listening situation the source decorrelation can be observed in an extended area in a larger listening area. Binaural listening depends on a constant listening position with only two loudspeakers available to represent the entire scene. The sum of several decorrelated sources convolved with different HRTF filters destroys the binaural sensation thus the maximum source size might lose its localization and externalization cues due to the superposition of decorrelated sources.

Because decorrelation is applied on a phantom source level (as opposed to applying a filter per speaker) the rotation of a fully distributed source is still perceivable, albeit its position is not. Yet, in the authors' opinion, this can reflect reality when considering a that a large source can be considered a construction of many smaller decorrelated sub-sources [2].

Considering that most loudspeaker arrays are more hemispherical than spherical for practical reasons, we recommend constructing any target configuration in such a way to give preference to sources moving directly overhead. In the tetrahedron case, we found results to be more convincing when the tetrahedron is rotated around the normal vector so that one decorrelated copy is rotated above the listener, if the original source's position is level, i.e. it's elevation is around $\frac{\pi}{2}$ (see figure 1(a)). Because Ambisonic sources exhibit a natural source size, the lower, lateral copies would be adequately rendered along the lower edge of a hemispherical layout.

The amount of modification on phase range and phase offset of the tested filters affects the perceived source size, listening area, frequency content and the perceived loudness. Not applying any phase offset to the random phase filters prevents decorrelation to a great degree. Increased offset distance between filters, though, also increases comb filtering effects if copies are close spatially. The range of decorrelation per filter reduces or increases spread sensation proportionally. In combination, a low decorrelation range but high offset causes prominent comb filtering effects, suggesting that a simple phase offset is not desirable with this approach. No difference was perceived between filters of either 128 or 512 taps. More lengths are needed to be tested.

Our tests also demonstrated that frequency dependent phase modification results in different effects on the overall source size perception. As previously mentioned, the phantom sources are filtered with different types of filters offering constant modification range and/or offset. Testing with monotonously increasing and decreasing filters, i.e. the lower frequencies are permitted to randomly shift in phase with larger values than higher frequencies and vice versa, instead of filters with a constant range and offset, we found out that small shifts in low frequencies keeps the low frequency image of the source intact while the higher frequencies are decorrelated equally. Small shifts in high frequencies do not seem to effect the overall size perception if the shift in low frequencies are still high. This is due to the differing wavelength of the frequencies. Small changes in listening position in a multichannel loudspeaker setup already cause drastic differences in phase for high frequencies with relatively small wavelengths. Low frequencies with relatively large wavelengths are more stable in their phase relationship in this regard. Because of this, decorrelation modifications are more important for low frequencies than higher ones.

Comb filtering can be reduced by introducing curves to the decorrelated source volume and decorrelated source position as discussed in section 2.4. This makes the spread behavior non-linear, as both volume and position changes are now skewed. This skewed spread behavior can be counteracted by also introducing a curve to the Ambisonics order reduction, reducing the order faster using its source widening effect earlier. This way, we combine the weaker spread sensation of the order reduction first and introduce the decorrelated sources later.

4 Summary

A new method for controlling the apparent source width applicable for both phantom sources and post-production is presented. The method combines the approach of reducing the ambisonic order of an input signal taking advantage of the inherent larger apparent source width of lower orders with a distribution of decorrelated copies around the listening space. The distribution is chosen to minimize comb filtering and use as few decorrelation copies as possible, while ensuring to cover the entire periphonic listening area at maximum desired source extent. Several decorrelation filters were constructed, compared and discussed. Additionally, the method was compared using different order reductions and distributions of the decorrelated copies. Overall, the method has shown to produce stable images of distributed sources in multichannel loudspeaker setups, with an impression of a source gradually increasing in size with increasing source size parameter α.

Preliminary tests were first necessary to reduce the amount of variables before formal testing can begin. Particularly in terms of filter construction, it was necessary to reduce the options available to the most effective first. Formal listening tests with trained and untrained listeners to evaluate the methods' effectiveness are subject to future work. Moreover, the evaluation of the correct combination between Ambisonic order reduction and decorrelated copy introduction is yet to be evaluated perceptually.

5 Acknowledgement

The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 732130 – BINCI project.

References

[1] Cabrera, A., *Control of source width in multichannel reproduction through sinusoidal modeling*, Ph.D. thesis, Queen's University Belfast, Belfast, U.K., 2012.

[2] Potard, G. and Burnett, I., "Decorrelation techniques for the rendering of apparent sound source width in 3D audio displays," in *Proc. Int. Conf. on Digital Audio Effects (DAFx'04)*, 2004.

[3] Weger, M., Marentakis, G., and Höldrich, R., "Auditory perception of spatial extent in the horizontal and vertical plane," in *Proc of the 19th International Conference on Digital Audio Effects (DAFx-16), Brno, Czech Republic*, 2016.

[4] Carpentier, T., "Ambisonic Spatial Blur," in *Audio Engineering Society Convention 142*, 2017.

[5] Cabrera, D. and Tilley, S., "Parameters for auditory display of height and size," in *Proceedings of the 2003 International Conference on Auditory Display*, Georgia Institute of Technology, Boston, MA, USA, 2003.

[6] Pulkki, V., "Uniform spreading of amplitude panned virtual sources," in *Applications of Signal Processing to Audio and Acoustics, 1999 IEEE Workshop on*, pp. 187–190, IEEE, 1999.

[7] Pihlajamäki, T., Santala, O., and Pulkki, V., "Synthesis of spatially extended virtual source with time-frequency decomposition of mono signals," *Journal of the Audio Engineering Society*, 62(7/8), pp. 467–484, 2014.

[8] Kendall, G. S., "The decorrelation of audio signals and its impact on spatial imagery," *Computer Music Journal*, 19(4), pp. 71–87, 1995.

[9] Bouéri, M. and Kyriakakis, C., "Audio signal decorrelation based on a critical band approach," in *Audio Engineering Society Convention 117*, Audio Engineering Society, 2004.

[10] Zotter, F., Frank, M., Kronlachner, M., and Choi, J.-W., "Efficient phantom source widening and diffuseness in ambisonics," *10.14279/depositonce-4103*, 2014.

[11] Gerzon, M. A., "General metatheory of auditory localisation," in *Audio Engineering Society Convention 92*, Audio Engineering Society, 1992.

[12] Schmele, T., García-Garzón, D., Sayın, U., Scaini, D., and Arteaga, D., "Layout remapping tool for multichannel audio productions," in *Audio Engineering Society Convention 134*, Audio Engineering Society, 2013.

[13] Daniel, J., *Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia*, Ph.D. thesis, University of Paris VI, France, 2000.

[14] Nishimura, T. and Matsumoto, M., "Mersenne Twister: A 623-dimensionally equidistributed uniform pseurandom number generator," in *ACM Transactions on Modeling and Computer Simulations: Special Issue on Uniform Random Number Generation*, 1998.