24th INTERNATIONAL CONGRESS ON SOUND AND VIBRATION
23–27 July 2017, London

ICSV24
23-27JULY 2017
LONDON CALLING

AN OPTIMIZATION APPROACH TO CONTROL SOUND SOURCE SPREAD WITH MULTICHANNEL AMPLITUDE PANNING

Andreas Franck
Filippo Maria Fazi
Eric Hamdan
Institute of Sound and Vibration Research,University of Southampton, Highfield, Southampton, UK
email: andreas.franck@soton.ac.uk, filippo.fazi@soton.ac.uk, eric.hamdan@soton.ac.uk

The perceived spread (or width) of a sound source is an important attribute of an audio object that should be controlled by a multi-channel audio reproduction system. It is desirable to either synthesize a plausible spread of a sound object or to maintain a constant spread when the virtual source moves. Existing sound spreading techniques such as Multiple Direction Amplitude Panning (MDAP) or time-frequency decomposition typically do not account for the specific loudspeaker arrangement and the inherent source spread generated by this layout. In this paper we propose an optimization-based sound field control approach, termed the ℓ_1/ℓ_2 method, to adjust the spread of a sound source. To this end we use the velocity vector magnitude to quantify the desired source spread. Based on the equivalence between amplitude panning and the maximization of this vector, the control of the source spread is formulated as a convex optimization problem. We show how the velocity objective function and additional constraints affect the resulting loudspeaker gain distribution. The proposed approach can be integrated into amplitude panning systems, and allows for either a position-independent spread of moving sources or for a smooth and continuous control of the spread parameter.

Keywords: Sound reproduction, source spread, source extent, object-based audio, amplitude panning

1. Introduction

Many real-world sound sources are characterized not only by their audio content and their perceived location, but also by an extent, also termed spread or width, that represents the perceived physical dimension of the sound source [6]. In order to recreate plausible auditory scenes, sound reproduction systems should therefore aim to recreate this spread. This holds true in particular for object-based audio systems, where the spread may form an attribute of an audio object that can be transmitted and manipulated as part of the metadata of the object. In addition, sound reproduction methods using discrete loudspeaker setups, for example such as Vector Base Amplitude Panning (VBAP) [1] produce an inherent spread that depends on the loudspeaker density and the source position relative to the loudspeakers. This causes fluctuations of the perceived spread if the source moves.

For use with VBAP, Pulkki [2] proposed the Multiple Direction Amplitude Panning technique (MDAP) which synthesizes a spread audio object by a group of coherent VBAP sources. Recently, this technique has been included in the reference renderer of the MPEG-H [3] standard which features an angular spread parameter. The influence of the interaural cross-correlation (IACC) [4] is utilized in reproduction techniques based on decorrelation, e.g., [5, 6, 7], including recent methods based on time-frequency decomposition [8].

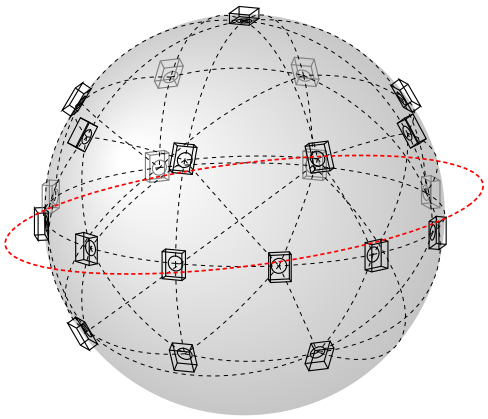


Figure 1: Example loudspeaker setup (ITU-R BS.2051 System H) with VBAP triangulation. The red circle represents the source trajectory used in the evaluation section.

Notwithstanding the use of a decorrelation technique, the selection and the driving gains of the active loudspeakers is a crucial aspect in the reproduction of an audio object with a given position and spread. Most existing methods either omit this step and assume a fixed selection, or, as in case of MDAP, use a simple heuristic approach that is not directly linked to the desired spread value. In addition, most selection schemes, with the exception of MDAP, are difficult to extend to 3D loudspeaker setups.

In this paper we investigate the use of optimization techniques to determine both the active loudspeakers and their gains in order to achieve a desired spread of a reproduced audio object. This approach is motivated by the relation between sparsity-enforcing ℓ_1 norm optimal panning and VBAP [9]. In particular, this builds upon the relation between the ℓ_1 norm and velocity or energy vector magnitude measures used to quantify source spread. We propose objective functions to calculate panning gain vectors, and evaluate the properties of the resulting sound events.

2. Panning Sound Reproduction Techniques

In this section we introduce VBAP as one form of amplitude panning, its relation to sparse, ℓ_1-optimal panning, and objective and subjective measures for the spread of sound sources.

2.1 Amplitude Panning and VBAP

Amplitude panning techniques create phantom images in the intended direction of an audio object. To this end, the audio signal is replayed through a subset of the loudspeakers of a discrete loudspeaker setup, each weighted by an individual real-valued panning gain. Vector Base Amplitude Panning (VBAP) [1] extends amplitude panning laws to 3D using vector notation. The direction of the virtual source and the loudspeakers positions are represented by unit vectors $\mathbf{p} = [x_p \; y_p \; z_p]^T$ and $\mathbf{l}_l = [x_l \; y_l \; z_l]^T$, respectively, corresponding to points on the unit sphere. The loudspeaker setup is expressed in matrix form as

$$\mathbf{L} = \begin{bmatrix} \mathbf{l}_1 & \mathbf{l}_2 & \cdots & \mathbf{l}_L \end{bmatrix} , \tag{1}$$

where L is the number of loudspeakers. The unit sphere is partitioned into a set of nonoverlapping loudspeaker triangles, typically using a Delaunay triangulation, e.g., [10]. An example of a multi-loudspeaker setup and of the corresponding triangulation is shown in Figure 1. This 22-channel setup is standardized as System H in ITU-R BS.2051 [11] and used throughout this paper.

To calculate the panning gains, a first step determines the triangle of active loudspeakers that contains the source direction $\mathbf{p}$. These loudspeakers denoted by the indices i, j, and k. In a second

step the panning gains for the active loudspeakers are determined as

$$\begin{bmatrix} g_i & g_j & g_k \end{bmatrix}^T = \begin{bmatrix} \mathbf{l}_i & \mathbf{l}_j & \mathbf{l}_k \end{bmatrix}^{-1} \mathbf{p}\,. \tag{2}$$

These gains are normalized to achieve a constant sound power level at the listener position [1]. By setting the gains of the other loudspeakers to zero, an overall gain vector $\mathbf{g} = [g_1\ g_2\ \cdots\ g_L]$ is formed.

2.2 Properties of VBAP

In addition to a low computational complexity, VBAP exhibits several properties that make it well-suited to practical sound reproduction. Firstly, it enables high timbral and localization quality, albeit in a small listening area. Outside this sweet spot, localization quality degrades gracefully by collapsing towards the closest active loudspeaker while preserving sound quality. Most of these advantages can be directly linked to properties of the VBAP algorithm:

Accurate particle velocity direction The direction of the particle velocity vector at the listening position matches that of the virtual source $\mathbf{p}$. This follows directly from the gain calculation (eq. (2)). Thus, VBAP ensures good localization at low frequencies (up to $\approx 700\,\mathrm{Hz}$) [12, 1, 13].

Sparsity Due to the geometric triangulation approach, a virtual sound source is reproduced by at most three loudspeakers for a 3D setup.

Locality Also following from the geometric approach, only loudspeakers in the vicinity of virtual source position are active.

Nonnegativity By construction, the gains of the active loudspeakers are positive. That means that sound quality degradations in the sweet spot due to destructive interference are substantially reduced, at least for low to mid frequencies.

Uniqueness Only a single valid VBAP solution exists if the Delaunay triangulation is unique.

One objective of this paper is to investigate whether these properties can be retained in a reproduction method for spread sound sources.

2.3 Sparse, ℓ_1-Optimal Amplitude Panning

In a recent article [9], authors of the present paper established a relation between sparse, ℓ_1-optimal amplitude panning and VBAP. Specifically, VBAP is identical to the (global) convex optimization problem

$$\underset{\mathbf{g}}{\operatorname{argmin}} \ \|\mathbf{g}\|_1 \tag{3a}$$

$$\text{subject to} \quad \mathbf{L}\mathbf{g} = \mathbf{p} \tag{3b}$$

$$\mathbf{g} \geq 0\,. \tag{3c}$$

if a Delaunay triangulation is used for the VBAP triangulation step. In this way, the sparsity-promoting nature of the ℓ_1 norm is directly linked to the geometric construction underlying VBAP.

2.4 Metrics to Evaluate Sound Source Spread

Measures for the perceived width of a sound source can be distinguished into objective and perceptual metrics. For amplitude panning techniques, the velocity and energy vectors (e.g., [12, 13, 10]) are widely used to describe perceived direction and spread

$$\mathbf{r}_\mathrm{v} = \hat{\mathbf{r}}_\mathrm{v} r_v = \frac{\sum_{i=1}^{L} \mathbf{l}_i g_i}{\sum_{i=1}^{L} g_i} \tag{4a}$$

$$\mathbf{r}_\mathrm{e} = \hat{\mathbf{r}}_\mathrm{e} r_e = \frac{\sum_{i=1}^{L} \mathbf{l}_i g_i^2}{\sum_{i=1}^{L} g_i^2}\,. \tag{4b}$$

The velocity vector $\mathbf{r}_\mathrm{v}$ is mainly useful for lower frequencies ($\leq 700\,\mathrm{Hz}$), while the energy vector $\mathbf{r}_\mathrm{e}$ described localization at higher frequencies. Both measures can be separated into a unit direction vector $\hat{\mathbf{r}}_{\{\mathrm{v|e}\}}$ and magnitude $r_{\{\mathrm{v|e}\}}$. Low values of $r_{\{\mathrm{v|e}\}}$ denote wide, spread sound events. Listening tests reported in [14] suggest that the energy vector magnitude r_e is a good predictor for source width. Both vector measures are purely based on loudspeaker positions and gains, and incorporate neither room acoustics nor properties of the loudspeaker signals such as decorrelation.

Optimization problem (3) implies that VBAP is equivalent to maximizing the velocity vector magnitude r_v while preserving the correct particle velocity direction [9]. This uses the fact that the ℓ_1 norm is identical to the sum of the elements of $\mathbf{g}$ and to the reciprocal of r_v in case of nonnegativity

$$\sum g_i = \|\mathbf{g}\|_1 = 1/r_\mathrm{v} \quad \text{if } g_i \geq 0\,,\, i = 1 \ldots L\,. \tag{5}$$

The second equality assumes that constraint (3b) holds. Thus, VBAP minimizes the low-frequency spread for a given source direction, as described qualitatively in [1, 2].

Measures to estimate the subjectively perceived source spread include the lateral fraction (LF) and the inter-aural cross-correlation (IACC) [4]. Their use with panning techniques has been evaluated in [14], which recommends the IACC_{E3} variant as a good estimator of the perceived spread. IACC_{E3} averages the maximum of the cross-correlation functions between the binaural ear signals in three octave bands centered at 500 Hz, 1 kHz, and 2 kHz. In this way, IACC includes the properties of both the listening environment and the loudspeaker signals.

3. Source Spreading Techniques

In this section we briefly review techniques for creating source spread with amplitude panning and propose a novel approach based on convex optimization, termed the ℓ_1/ℓ_2 method.

3.1 Multiple Direction Amplitude Panning (MDAP)

MDAP was proposed by Pulkki [2] as an extension to VBAP to achieve more uniform or extended sound sources. To this end it creates a set of additional source directions, driven coherently by the same source signal, around the original direction $\mathbf{p}$. Recently MDAP has been adopted as the spreading technique in the MPEG-H standard [15, 3]. The reference implementation of this standard allows features an angular spread parameter α, $0^\circ \leq \alpha \leq 180^\circ$ and creates a fixed set of 18 directions for each spread source.

3.2 Decorrelation-Based Methods

Approaches to create perceived sound source spread by decorrelating the loudspeaker signals date back to pseudo-stereo techniques [16] and evolved into more sophisticated decorrelation techniques, e.g., [5, 6, 7]. As an advantage over coherent reproduction techniques as MDAP, they directly affect the IACC. According to [7], this is a necessary requirement for realistic source spread. More recently, [8] presents a spreading method that decorrelates a monophonic signal by time-frequency processing.

Notwithstanding the use of a decorrelation technique, the selection of the loudspeakers as well as the gains used for driving these loudspeakers remains a crucial step in spreading algorithms.

3.3 Proposed ℓ_1/ℓ_2 method

In this paper we propose a novel method to determine panning gain vectors to synthesize virtual sound sources with controllable spread. Thus, it is applicable both for coherent reproduction techniques as MDAP as well as in combination with decorrelation.

The representation of VBAP as an optimization problem (3) and the equivalence between minimizing the ℓ_1 norm of $\mathbf{g}$ and maximizing the velocity vector magnitude r_v forms the starting point

for this method. Instead of maximizing the velocity vector magnitude, the optimization problem is altered such that the desired spread value is added as an equality constraint based on the desired velocity vector magnitude $\tilde{r}_\mathrm{v}$. Assuming a unit norm of the resulting velocity vector (3b), this implies

$$\|\mathbf{g}\|_1 = 1/\tilde{r}_\mathrm{v}\,. \tag{6}$$

However, this additional constraint fixes the value of the objective function (3a) of the original optimization problem to a constant value, thus rendering it ineffective. The resulting system is typically underdetermined and therefore has a multitude of solutions which might activate very different sets of loudspeakers.

To preserve solution uniqueness as well as the advantageous properties of VBAP-like panning algorithms, a new objective function has to be found. In this paper we propose using a ℓ_2 norm minimization objective on the loudspeaker gain vector. Because this objective results in the least-squares solution subject to a ℓ_1 equality constraint on $\mathbf{g}$, we refer to it as the ℓ_1/ℓ_2 method in the following. The complete optimization problem reads

$$\underset{\mathbf{g}}{\operatorname{argmin}} \ \|\mathbf{g}\|_2 \tag{7a}$$

$$\text{subject to} \quad \mathbf{L}\mathbf{g} = \mathbf{p} \tag{7b}$$

$$\|\mathbf{g}\|_1 = 1/\tilde{r}_\mathrm{v} \tag{7c}$$

$$\mathbf{g} \geq 0\,. \tag{7d}$$

In other words, this optimization problem yields the minimum-energy panning function that preserves the correct particle velocity direction and velocity vector magnitude while enforcing gain nonnegativity. This problem is solvable only if the desired velocity vector magnitude $\tilde{r}_\mathrm{v} = 1/\|\mathbf{g}\|_1$ does not exceed the maximum value corresponding to the minimum $\|\mathbf{g}\|_1$ defined by the VBAP solution (3).

Figure 2 illustrates the effects of this objective function and compares it to the MDAP algorithm as specified in the MPEG-H standard [15]. To this end, the ℓ_1/ℓ_2 has been implemented in the convex optimization modeling toolkit CVX [17, 18]. For both algorithms, a spread sound source with position $\mathbf{p} = (40°, 15°)$ and a spread parameter $\alpha = 20°$, as defined by MPEG-H [15], is synthesized. In case of MDAP, this leads to nine active loudspeakers, which are located up to $\approx 75°$ away from the desired source direction. In contrast, the ℓ_1/ℓ_2 method uses four loudspeakers that have a maximum angular distance of $\approx 39.6°$ to the desired object direction $\mathbf{p}$. This demonstrates that, in this example, the proposed technique apparently preserves advantageous features of VBAP, in particular the sparsity and locality of active loudspeakers.

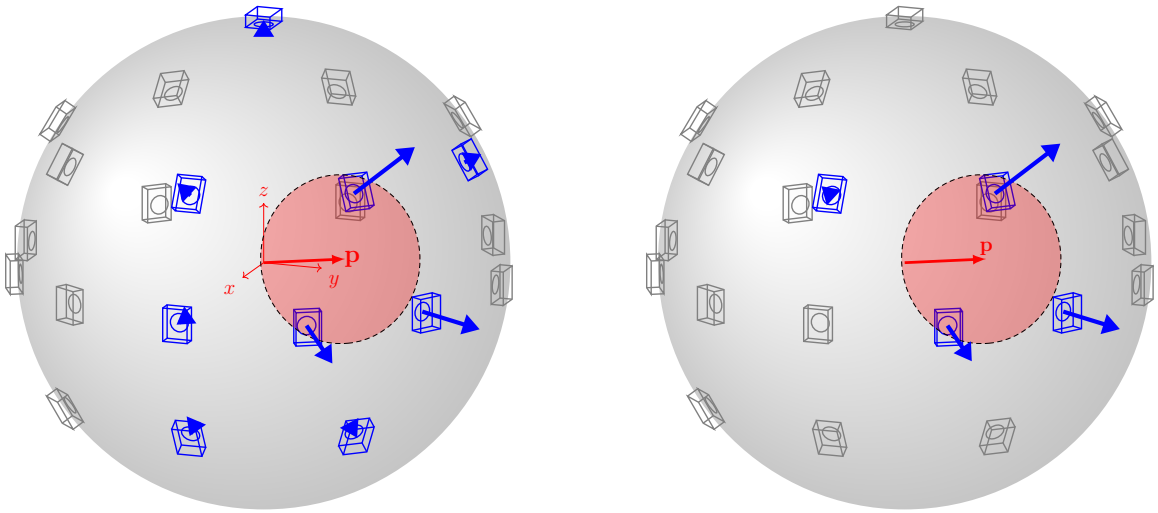
4. Evaluation

In this section we evaluate the considered spreading techniques using objective and subjective measures. To this end, we synthesize a sound source moving on a circular trajectory on a horizontal plane that is tilted by $7.5°$ with respect to the x axis. Both the setup and the source trajectory are shown in Figure 1. A value of $\alpha = 55°$ is selected as the desired source spread parameter. The corresponding velocity vector magnitude $r_\mathrm{v} \approx 0.787$ is determined by assuming a homogeneous source strength over this circular distribution, resulting in

$$r_\mathrm{v} = \tfrac{1}{2}\left[\cos\alpha + 1\right]\,. \tag{8}$$

Note that this definition differs from the "spherical cap" analogy used in [10].

As described in Sec. 2.4, the velocity and energy vector magnitudes can be used as metrics for the source spread. In Figures 3(a) and 3(b) they are displayed as functions of the source azimuth for the chosen source trajectory. The velocity vector magnitude is shown for MDAP, the proposed ℓ_1/ℓ_2 method and, for comparison, the corresponding VBAP panning gains of a non-spread source.



(a) MDAP (b) ℓ_1/ℓ_2 method

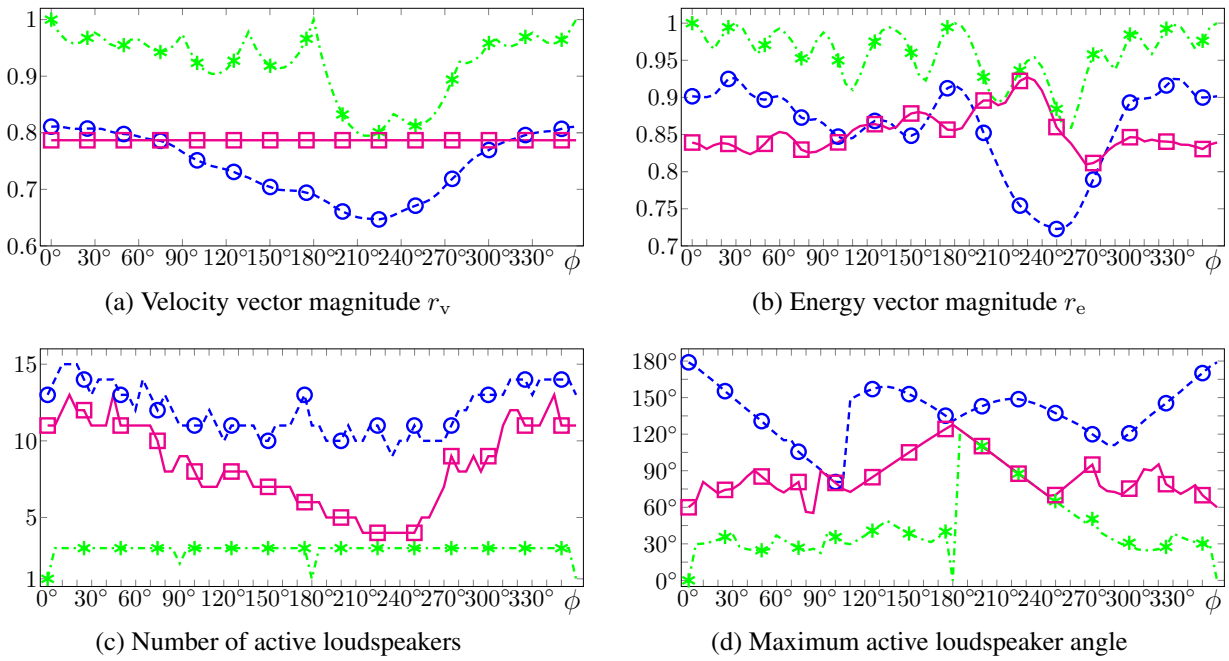
Figure 2: Loudspeaker gain distributions for source position $\mathbf{p} = (40°, 15°)$, spread $\alpha = 20°$. The red disks visualize the desired spread. Active loudspeakers and their gains are displayed in blue.

It is observed that r_v varies considerably in case of MDAP, with lower values at positions where the corresponding VBAP panning also has a low r_v. Not surprisingly, the value for the proposed ℓ_1/ℓ_2 method is constant at $r_\mathrm{v} \approx 0.787$, because this value is enforced by the optimization constraint (eq. (7c)). The corresponding energy vector magnitudes are shown in Fig. 3(b). Unlike r_v, this measure is not used in the objective function of the ℓ_1/ℓ_2 method. Nonetheless, this method reduces the variation of r_e significantly, which can be quantified by the standard deviation $\sigma_{\ell_1/\ell_2} = 0.027$ compared to the values of $\sigma_\mathrm{VBAP} = 0.035$ and $\sigma_\mathrm{MDAP} = 0.058$ for VBAP and MDAP, respectively. According to [14], this implies a more uniform source spread for the proposed technique.

Figures 3(c) and 3(d) show the number of loudspeakers with nonzero gains and the maximum angular distance between the source center and a loudspeaker with nonzero gain. Both measures show significantly lower numbers for the proposed technique than for MDAP. This implies that the combined ℓ_1/ℓ_2 criterion, to some extent, preserves the favorable sparsity and locality properties that underlie amplitude panning techniques as VBAP.

To obtain indications about the perceived source spread, the interaural cross correlation-coefficient (IACC), specifically the IACC_{E3} measure, is evaluated for the same source trajectory as for the objective performance measures above. The binaural ear signals are generated by applying the BRIR dataset [19] of a 22.2 multichannel loudspeaker system installed in a listening room that conforms to ITU-R BS.1116. To investigate the effect of decorrelated loudspeaker signals, a decorrelation using a bank of random-phase FIR allpass decorrelators (length 512 taps), e.g., [5], is optionally applied.

The resulting IACC values are shown in Figure 4. It is observed that the methods without decorrelation (VBAP, MDAP, and ℓ_1/ℓ_2) do not achieve a significantly lowered IACC for most positions on the trajectory. This coincides with observations, e.g., [7, 8], that a coherent reproduction over multiple loudspeakers is not perceived as a spatially spread sound source. In contrast, both spreading techniques using decorrelation achieve significantly reduced IACC, with mean values of ≈ 0.46 and ≈ 0.36 for MDAP + decorrelation and ℓ_1/ℓ_2 + decorrelation, respectively. However, the fluctuation of the proposed method is significantly lower, which is confirmed by the standard deviations $\sigma_{\ell_1/\ell_2+\mathrm{decorr}} \approx 0.053$ compared to $\sigma_{\mathrm{MDAP}+\mathrm{decorr}} \approx 0.105$. In particular, the ℓ_1/ℓ_2 approach does not show the tendency of all other methods towards higher IACC values if the source position is close to a loudspeaker. In the considered loudspeaker layout, this corresponds to azimuth angles of about $0°$, $30°$, $180°$, and $360°$. It is noted that the proposed method achieves these IACC values using a smaller and more localizes subset of active loudspeakers, as shown in Figures 3(c) and 3(d). In combination,



(a) Velocity vector magnitude r_v

(b) Energy vector magnitude r_e

(c) Number of active loudspeakers

(d) Maximum active loudspeaker angle

Figure 3: Objective performance measures. -*-: VBAP, -o-: MDAP, -□-: Proposed ℓ_1/ℓ_2 method.

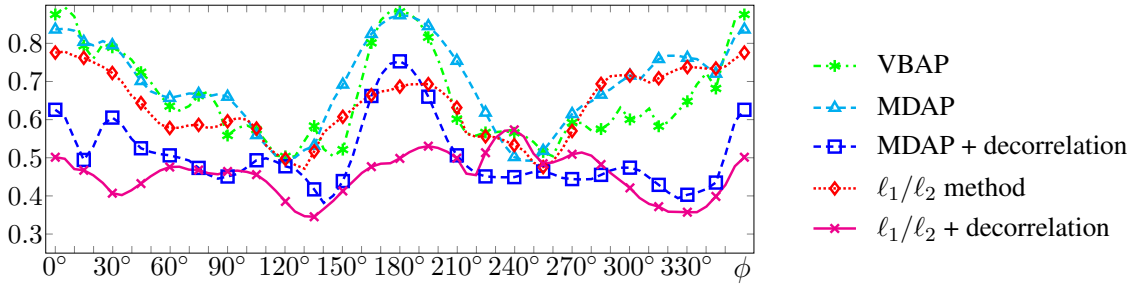


Figure 4: Interaural cross correlation coefficients simulated for ITU-R BS.1116 listening room [19].

these performance measures suggest that the ℓ_1/ℓ_2 method enables a more consistent, controllable spread for a wide range of source positions.

5. Conclusion

In this paper we investigated the application of convex optimization techniques to synthesize sound sources with a controllable spread by means of amplitude panning techniques. Specifically, we considered means to determine the set of active loudspeakers and their gains that can be used both with coherent and decorrelated loudspeaker signals. The results show that the proposed ℓ_1/ℓ_2 technique allows a significantly better control of the source spread than existing methods as MDAP, both in terms of objective measures as the velocity vector and energy magnitudes as well as psychoacoustic measures as IACC. In addition, advantageous properties of panning technique, in particular sparsity and locality of active loudspeakers, are better preserved by the proposed optimization criterion. Further research will focus on perceptive evaluation in real-time reproduction systems, and on efficient methods to determine or approximate such optimal loudspeaker gain distributions.

Acknowledgment

This work was supported by the EPSRC Programme Grant S3A: Future Spatial Audio for an Immersive Listener Experience at Home (EP/L000539/1) and the BBC as part of the BBC Audio Research Partnership.

REFERENCES

1. Pulkki, V. Virtual sound source positioning using vector base amplitude panning, *J. Audio Eng. Soc.*, **45** (6), 456–466, (1997).
2. Pulkki, V. Uniform spreading of amplitude panned virtual sources, *Proc. IEEE Workshop Applicat. Signal Process. Audio Acoust.*, New Paltz, NY, USA, Oct., pp. 187–190, (1999).
3. Herre, J., Hilpert, J., Kuntz, A. and Plogsties, J. MPEG-H 3D audio - the new standard for coding of immersive spatial audio, *IEEE J. Sel. Topics Signal Process.*, **9** (5), (2015).
4. Hidaka, T., Beranek, L. L. and Okano, T. Interaural cross-correlation, lateral fraction, and low- and high-frequency sound levels as measures of acoustical quality in concert halls, *J. Acoust. Soc. Am.*, **98** (2), 988–1007, (1995).
5. Kendall, G. S. The decorrelation of audio signals and its impact on spatial imagery, *Computer Music Journal*, **19** (4), 71–87, (1995).
6. Potard, G. and Burnett, I. Decorrelation techniques for the rendering of apparent sound source width in 3D audio displays, *Proc. 7th Int. Conf. Digital Audio Effects (DAFx'04)*, Naples, Italy, Oct., pp. 280–284, (2004).
7. Potard, G., *3D-Audio Object Oriented Coding*, Ph.D. thesis, University of Wollongong, Wollongong, Australia, (2006).
8. Pihlajamäki, T., Santala, O. and Pulkki, V. Synthesis of spatially extended virtual sources with time-frequency decomposition of mono signals, *J. Audio Eng. Soc.*, **62** (7/8), 467–484, (2014).
9. Franck, A., Wang, W. and Fazi, F. M. Sparse, ℓ_1-optimal multi-loudspeaker panning and its relation to vector base amplitude panning, *IEEE/ACM Trans. Audio, Speech, Language Process.*, DOI 10.1109/TASLP.2017.2674975, (2017).
10. Zotter, F. and Frank, M. All-round Ambisonic panning and decoding, *J. Audio Eng. Soc.*, **60** (10), 807–820, (2012).
11. ITU. ITU-R BS.2051-0 — Advanced Sound System for Programme Production, (2014).
12. Gerzon, M. A. Panpot laws for multispeaker stereo, *AES 92th Convention*, Vienna, Austria, Mar., (1992).
13. Jot, J.-M., Larcher, V. and Pernaux, J.-M. A comparative study of 3-D audio encoding and rendering techniques, *Proc. AES 16th Int. Conf.*, Rovaniemi, Finland, Mar., (1999).
14. Frank, M. and Zotter, F. Simple technical prediction of phantom source widening, *Proc. AIA-DAGA 2013 Conf. Acoustics*, Meran, Italy, Mar., (2013).
15. ISO/IEC, Information technology—High efficiency coding and media delivery in heterogeneous environments—Part 3: 3D Audio. ISO/IEC Standard ISO/IEC 23008-3:2015, (2015).
16. Schroeder, M. R. An artificial stereophonic effect obtained from a single audio signal, *J. Audio Eng. Soc.*, **6** (2), 74–79, (1958).
17. Grant, M. and Boyd, S., (2008), Graph implementations for nonsmooth convex programs. Blondel, V., Boyd, S. and Kimura, H. (Eds.), *Recent Advances in Learning and Control*, pp. 95–110, Springer.
18. Grant, M. and Boyd, S., (2016), *CVX: Matlab Software for Disciplined Convex Programming, version 2.1*. `http://cvxr.com/cvx`.
19. Francombe, J., (2016), *IoSR Listening Room Multichannel BRIR Dataset*. DOI 10.15126/surrey-data.00813511.