PARAMETRIC BINAURAL RENDERING UTILIZING COMPACT MICROPHONE ARRAYS

Symeon Delikaris-Manias, Juha Vilkamo and Ville Pulkki

Aalto University, School of Electrical Engineering,
Department of Signal Processing and Acoustics, Espoo, FI-00076, Finland

ABSTRACT

The spatial capture patterns according to head-related transfer-functions (HRTFs) can be approximated using linear beamforming techniques. However, assuming a fixed spatial aliasing frequency, with reduction of the number of sensors and thus the array size, the linear approach leads to an excessive amplification of the microphone noise, unless the beam patterns are made broader than determined by the HRTFs. An adaptive technique is proposed that builds upon the assumption that the binaural perception is largely determined by a set of short-time inter-aural parameters in frequency bands. The parameters are estimated from the noisy HRTF beam pattern signals as a function of time and frequency. As a result of temporal averaging, the effect of the noise is mitigated while the perceptual spatial information is preserved. Signals with higher SNR from broader patterns are adaptively processed to obtain the parameters from the estimation stage by means of least-squares optimized mixing and decorrelation. Listening tests confirmed the perceptual benefit of the proposed approach with respect to linear techniques.

Index Terms— binaural rendering, microphone arrays, time-frequency processing, beamforming.

1. INTRODUCTION

Microphone array signal processing techniques are commonly utilized in beamforming applications. One of the applications is approximating the directional characteristics of head-related transfer functions (HRTFs) for binaural reproduction of a sound scene. The beam patterns can be designed in the sensor domain or by utilizing the spherical harmonic framework [1, 2, 3, 4]. The rendering accuracy depends on the number of microphones as well as the geometry and the size of the microphone array. At high frequencies, especially for sparse arrays, the reproduction is also affected by spatial aliasing [5]. For compact arrays, the linear beamforming techniques approximating HRTF patterns accurately are subject to excessive microphone self noise amplification.

In this paper, an adaptive method is proposed that builds upon the assumption that the binaural perception is largely determined by a set of short-time stochastic properties in frequency bands, which are the inter-aural level difference (ILD), inter-aural phase difference (IPD), inter-aural coherence (IC), and the sum energy of the binaural frequency band signals. These properties are expressed in frequency bands by the binaural signal covariance matrix. In the proposed method, the covariance matrix of spatially selective beam patterns that are designed with respect to a target set of HRTFs are analyzed over time-frequency regions. Through the averaging process, the effect of the noise is mitigated in the estimates. Signals from broader, and less noisy, beam patterns are adaptively processed such that they obtain the covariance matrix determined at the analysis stage. By these means the signal quality of the broader beam patterns is combined with the spatial properties of the narrower beam patterns. Since the proposed method does not comprise any scheme for increasing the modal resolution of microphone arrays with a limited number of microphones, the effect of spatial aliasing is inherited. However, techniques that mitigate the effect of the reduced modal order [6] can readily be employed.

The research leading to these results has received funding from the Academy of Finland, the Aalto ELEC school and the Nokia Foundation.

2. INTERPOLATION OF HEAD RELATED TRANSFER FUNCTIONS

The proposed binaural beamforming technique is designed with respect to the left and right ear HRTFs $H_{\mathrm{l}}(\theta, \phi, k)$ and $H_{\mathrm{r}}(\theta, \phi, k)$, which are complex values as a function of source azimuth θ, elevation ϕ, and the frequency band index k. The amplitudes and phases of the HRTFs define the inter-aural level and phase differences, as well as the overall sum energy that is relevant especially for the localization within a cone of confusion (Fig. 1). The CIPIC head-related impulse response (HRIR) database [7] is employed in this study. Similar principles are applicable for processing other databases, for example, individualized HRIR measurements. The interpolation and equalization of the CIPIC HRIRs is described in the following, assuming diffuse field equalized headphones.

The perceptually relevant information in a pair of HRIRs is contained in their energies as a function of frequency and a single wide band inter-aural time-difference (ITD) parameter [8]. The ITD can be estimated as the difference of the median values of the group delays of the HRIR pair. The spectra

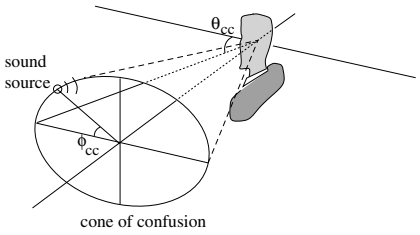

Fig. 1. The cone of confusion, and the spherical coordinate system ($\theta_{\rm cc}$,$\phi_{\rm cc}$) applied by the CIPIC HRIR database [7].

can be estimated by decomposing the HRIRs into frequency bands in which their energies are formulated. It is assumed that a sufficient frequency resolution for the energy estimation is determined by the Bark bands [9]. The complex-modulated hybrid QMF filter bank, applied throughout this manuscript and previously, e.g., in [10, 11, 12], approximates such a resolution.

Having parametrized the set of HRIRs to a set of wide band ITDs and energies as a function of frequency, the parameters at arbitrary data points can be obtained with the following steps by bilinear interpolation. The CIPIC database applies the spherical coordinate system ($\theta_{\rm cc}$,$\phi_{\rm cc}$) in Fig. 1. The position of a target data point (θ, ϕ) is first translated to this coordinate system, and ITD and energies are then linearly interpolated separately, first along the elevation axis, then azimuth. As a result, the energy and time-difference parameters $E_{\rm l}(\theta, \phi, k)$, $E_{\rm r}(\theta, \phi, k)$, and ${\rm ITD}(\theta, \phi)$ at the target data point are obtained. The ITD parameter is translated into an IPD parameter between $\pm\pi$ by

$$\mathrm{IPD}(\theta, \phi, k) = [(\mathrm{ITD}(\theta, \phi) \cdot 2\pi f_{\rm b}(k) + \pi) \mod 2\pi] - \pi, \quad (1)$$

where $f_{\rm b}(k)$ is the band center frequency. The interpolated HRTFs are then

$$\begin{aligned} H_{\rm l}(\theta, \phi, k) &= e^{i \cdot \mathrm{IPD}/2} \sqrt{E_{\rm l}(\theta, \phi, k)} G(k) \\ H_{\rm r}(\theta, \phi, k) &= e^{-i \cdot \mathrm{IPD}/2} \sqrt{E_{\rm r}(\theta, \phi, k)} G(k), \end{aligned} \quad (2)$$

where $G(k)$ is a diffuse-field equalizing gain that is formulated such that

$$\frac{1}{A} \sum_{a=0}^{A-1} \left[\|H_{\rm l}(\theta_a, \phi_a, k)\|^2 + \|H_{\rm r}(\theta_a, \phi_a, k)\|^2 \right] = 1, \quad (3)$$

for a set of data points (θ_a, ϕ_a) selected to be uniformly distributed across a sphere, where $a = 0...(A-1)$ and A is the total number of data points.

3. PROPOSED RENDERING ALGORITHM

The following formulation assumes time-frequency transformed signals. Any transform is applicable that allows independent processing of the frequency bands, such as the complex-modulated QMF bank [13] and the robust short-time Fourier processing techniques [14].

3.1. Rendering method

The proposed rendering method applies parametrized processing to the microphone signals $\mathbf{x}(k, l)$, where l is the time index, based on three sets of static beamforming weights: the analysis matrix $\mathbf{W}_{\rm a}(k)$ that approximates the HRTFs spatially accurately, being thus subject to excessive microphone self noise amplification, the synthesis matrix $\mathbf{W}_{\rm s}(k)$ that only loosely approximates the HRTFs and produces signals with a higher SNR, and the sum row vector $\mathbf{w}_{\rm o}^T(k)$ that approximates the sum spatial energy capture pattern of the left and right HRTFs with a single beam pattern. The indices (k, l) are omitted in the following for the brevity of notation.

The processing block diagram is shown in Fig. 2. The processing is performed in time-frequency areas in terms of covariance matrices, which contain information of the perceptually relevant binaural cues. The covariance matrix $\mathbf{C}_{\rm a}$ of the analysis signal $\mathbf{W}_{\rm a}\mathbf{x}$ is first estimated by $\mathbf{C}_{\rm a} = \mathrm{E}\left[(\mathbf{W}_{\rm a}\mathbf{x})(\mathbf{W}_{\rm a}\mathbf{x})^H\right]$, where $\mathrm{E}[\cdot]$ denotes the expectation. In a practical implementation, the expectation is approximated with a mean operator over the time-frequency interval. The target covariance matrix $\mathbf{C}_{\rm y}$ is then obtained by normalizing $\mathbf{C}_{\rm a}$ to have the estimated energy e_o of the sum signal $\mathbf{w}_{\rm o}^T\mathbf{x}$. The covariance matrix $\mathbf{C}_{\rm s}$ of the synthesis signal $\mathbf{W}_{\rm s}\mathbf{x}$ is also estimated. A least squares optimized mixing solution is formulated, as described in Section 3.3, to process the synthesis signal $\mathbf{W}_{\rm s}\mathbf{x}$ to have the target covariance matrix $\mathbf{C}_{\rm y}$.

3.2. Formulation of the weights matrices

The matrices $\mathbf{W}_{\rm a}$, $\mathbf{W}_{\rm s}$, and vector $\mathbf{w}_{\rm o}^T$ are derived with different static beamforming designs. These designs are based on conventional least-squares beamforming, in which the complex target HRTF beam patterns are approximated by utilizing the steering vectors $\mathbf{V}$ of the microphone array. The weights are calculated by

$$\mathbf{W}_{\rm a} = \mathbf{V}^{+}\mathbf{H}, \quad (4)$$

$$\mathbf{w}_{\rm o} = \mathbf{V}^{+}\mathbf{h}_o, \quad (5)$$

$$\mathbf{W}_{\rm s} = [\mathbf{V}^T\mathbf{V} + \beta\mathbf{I}]^{-1}\mathbf{V}^T\mathbf{H}, \quad (6)$$

where $[^{+}]$ indicates the Moore-Penrose inverse and $\mathbf{H}$ consists of the diffuse-field equalized HRTFs as defined in (2), for the same set of directions as the steering vectors. Vector $\mathbf{h}_o$ consists of the gains corresponding to the sum energy of the HRTFs, also for the same set of directions, and β is a regularization parameter. Hence, the weights for the analysis and sum patterns in (4) and (5) are approximated with a non-regularized Moore-Penrose solution, while for the synthesis patterns in (6) a Tikhonov regularization scheme is applied as in [15].

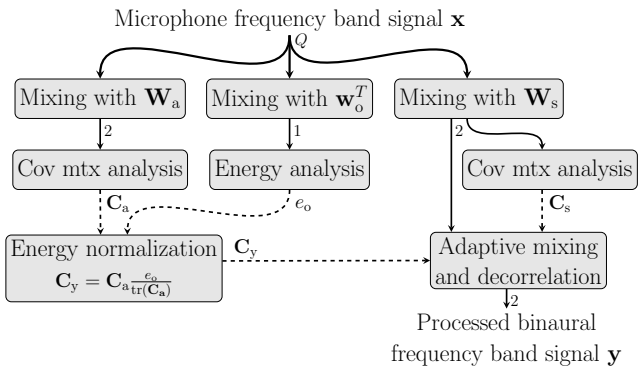

Fig. 2. Block diagram of the proposed method. The numbers denote the number of channels. Q is the number of microphones. The dashed lines denote parametric information at a lower sampling rate.

3.3. Adaptive mixing and decorrelation

At low frequencies ($f < 2700$ Hz), in which the human hearing is more sensitive to the inter-aural phase difference and coherence, the adaptive mixing block in Fig. 2 employs the technique proposed in [16]. The technique solves mixing matrices $\mathbf{M}$ and $\mathbf{M}_\mathrm{r}$ in a least-squares sense using an input-output relation

$$\mathbf{y} = \mathbf{M}(\mathbf{W}_\mathrm{s}\mathbf{x}) + \mathbf{M}_\mathrm{r}\mathrm{D}[(\mathbf{W}_\mathrm{s}\mathbf{x})], \tag{7}$$

where $\mathrm{D}[\cdot]$ is a decorrelating signal processing operation and $\mathbf{y}$ is the output signal. The matrix $\mathbf{M}$ is first solved with respect to a boundary condition that $\mathbf{y}$ obtains the target covariance matrix $\mathbf{C}_\mathrm{y}$. However, regularization is employed to avoid excessively large mixing coefficients, which could result in observable amplification of the microphone self noise, or other artifacts. The secondary mixing matrix $\mathbf{M}_\mathrm{r}$ is solved such that the decorrelated signals $\mathrm{D}[(\mathbf{W}_\mathrm{s}\mathbf{x})]$ are processed obtain a covariance matrix $\mathbf{C}_\mathrm{r}$ that is complementary to the effect of the regularization, so that

$$\mathbf{C}_\mathrm{y} = \mathbf{M}(\mathbf{W}_\mathrm{s}\mathbf{x})(\mathbf{W}_\mathrm{s}\mathbf{x})^H\mathbf{M}^H + \mathbf{C}_\mathrm{r}. \tag{8}$$

Eq. (8) requires that $\mathrm{E}\left[\mathrm{D}[(\mathbf{W}_\mathrm{s}\mathbf{x})](\mathbf{W}_\mathrm{s}\mathbf{x})^H\right] = \mathbf{0}$. Assuming first that $\mathbf{M}_\mathrm{r} = \mathbf{0}$, and setting the boundary condition that the output signal $\mathbf{y}$ must obtain $\mathbf{C}_\mathrm{y}$, the optimized $\mathbf{M}$ is formulated as

$$\arg\min_{\mathbf{M}} \mathrm{E}\left[\|\mathbf{G}(\mathbf{W}_\mathrm{s}\mathbf{x}) - \mathbf{M}(\mathbf{W}_\mathrm{s}\mathbf{x})\|^2\right], \tag{9}$$

where $\mathbf{G}$ is a matrix that equalizes the channel energies of $(\mathbf{W}_\mathrm{s}\mathbf{x})$ to those of $\mathbf{y}$. It is defined by

$$\mathbf{G} = \left(\mathrm{Diag}[\mathbf{C}_\mathrm{y}]\mathrm{Diag}[\mathbf{C}_\mathrm{s}]^{-1}\right)^{\frac{1}{2}}, \tag{10}$$

where $\mathrm{Diag}[\cdot]$ denotes an operation preserving only the diagonal of the matrix. Following the steps in [16], the solution to (9) with respect to the defined boundary condition is $\mathbf{M} = \mathbf{K}_\mathrm{y}\mathbf{V}\mathbf{U}^H\mathbf{K}_\mathrm{s}^{-1}$, where $\mathbf{K}_\mathrm{s}$ and $\mathbf{K}_\mathrm{y}$ are any decompositions fulfilling $\mathbf{C}_\mathrm{s} = \mathbf{K}_\mathrm{s}\mathbf{K}_\mathrm{s}^H$ and $\mathbf{C}_\mathrm{y} = \mathbf{K}_\mathrm{y}\mathbf{K}_\mathrm{y}^H$, and $\mathbf{V}$ and $\mathbf{U}$ are the unitary matrices from a singular value decomposition (SVD) with a relation $\mathbf{U}\mathbf{S}\mathbf{V}^H = \mathbf{K}_\mathrm{s}^H\mathbf{G}^H\mathbf{K}_\mathrm{y}$, and $\mathbf{S}$ is a non-negative diagonal matrix.

The inverse $\mathbf{K}_\mathrm{s}^{-1}$ is regularized using SVD and limiting the gains of the inverse of the diagonal matrix. Also other regularizations are applicable. Since the regularization affects the obtained target covariance matrix, a signal $\mathbf{M}_\mathrm{r}\mathrm{D}[(\mathbf{W}_\mathrm{s}\mathbf{x})]$ is synthesized with a complementary covariance matrix $\mathbf{C}_\mathrm{r}$ in (8). The decorrelated signal $\mathrm{D}[(\mathbf{W}_\mathrm{s}\mathbf{x})]$ has a covariance matrix $\mathrm{Diag}[\mathbf{C}_\mathrm{s}]$. The secondary mixing matrix $\mathbf{M}_\mathrm{r}$ is formulated with a corresponding procedure as with $\mathbf{M}$, however with $\mathbf{C}_\mathrm{r}$ as the target covariance matrix. An elaborated description of the techniques involved can be found in [16].

At high frequencies ($f \geq 2700$ Hz) the human hearing is less sensitive to the IPD and the coherence. Thus, at these frequencies only gain modulation is applied to synthesize the overall energy and the ILD by $\mathbf{M} = \mathbf{G}$ and $\mathbf{M}_\mathrm{r} = \mathbf{0}$.

4. EXPERIMENTAL VALIDATION

The perceptual quality provided by the proposed technique was evaluated with listening experiments. An offline Matlab software was implemented to process microphone recordings with the proposed method and the comparison beamforming techniques. A complex modulated 64-band QMF bank with cascaded sub-subband filters at the three lowest frequency bands was applied as the time-frequency transform. The frequency band signals were processed in frames of 32 QMF time indices. The covariance matrix was estimated in each frequency band with a rectangular window of 64 QMF time indices. The mixing matrices were formulated for each frame as described in Section 3.3, and linearly interpolated between the frames. The decorrelators were random delays in each frequency band and channel with the limits of 2-20 QMF time indices in the lowest frequency band, transitioning to 2-10 QMF time indices in the highest frequency band. An adaptive process for suppressing onsets and transients was applied prior to the decorrelators to avoid processing artifacts. The static beamforming comparison modes were the analysis beam pattern signals $\mathbf{W}_\mathrm{a}\mathbf{x}$ and the synthesis beam pattern signals $\mathbf{W}_\mathrm{s}\mathbf{x}$.

4.1. Reference sound scenes and reproduction modes

Three reference sound scenes were generated containing simultaneous sources and a simplified room effect. The sources were simulated as point sources in the horizontal plane in a free field. The room effect was simulated using 32 independent channels of reverberation, which were reproduced as free field point sources distributed at the vertices of a dodecahedron-icosahedron compound. The reverberation responses, which were applied as a convolution to the source

signals, were white Gaussian noise sequences modulated in frequency bands with exponentially decaying windows according to pre-defined reverberation times. The reverberation times ranged from 1.0 seconds in the lowest frequency band to 0.25 seconds in the highest frequency band.

Three sound scenes were synthesized: (Single) scene including a male talker at $90°$, (double) scene including female talker and classical guitar at $\pm 90°$, and (music) scene containing four instruments at $\pm 90°$ and $\pm 45°$. Front sources were not included since non-individual HRTFs provide poor binaural externalization at the front.

The (reference) mode was generated by applying the diffuse field equalized HRTFs to the set of point sources corresponding to the direct and reverberated sounds. The (mono) anchor mode was the sum of the point source signals without employing binaural processing. The other modes were generated by simulating a recording with a 10^{th} order uniform spherical microphone array of 3 cm radius and performing the binaural rendering. The (analysis), (synthesis), and (proposed) modes were processed using the Matlab test implementation, and were subject to a mean SNR of 65 dB. Furthermore, a (noiseless) mode was generated which was the same as (analysis) mode, but without noise except the low-level quantization noise that originated from the storing of the audio data in the 16-bit PCM format. For fair comparison, all reproduction modes were equalized to have in average the spectrum of the (reference) mode, since the (proposed) mode inherently has processes to ensure such a spectrum.

4.2. Subjective evaluation

A listening test was performed in an isolated listening booth using a multiple-stimulus test with a known reference that was included also as a hidden reference. The order of the reproduction modes was randomized, and their difference with respect to the reference was evaluated with sliders. The top of the slider (100 points) was denoted to mean that no difference is perceivable, and scores towards the low an increasing difference. The listeners were instructed to apply the scale broadly. The test was repeated with Stax SR-307 and Sennheiser HD-600 headphones with randomised order. Ten subjects participated to the test, all of which researchers in the field of audio, excluding the authors of this study.

The result data was analyzed using SPSS Statistics [17] with a three-way repeated measures analysis of variance with factors (reproduction_mode), (sound_scene), and (headphones). Mauchly's test revealed that the assumption of sphericity was violated with factor (sound_scene), and the Greenhouse-Geisser correction was applied. Significant effects were found with factor (reproduction_mode) ($p = 0.000$) and with interaction (reproduction_mode)*(sound_scene) ($p = 0.002$). The means and 95% confidence intervals of factor (reproduction_mode) across all sound scenes and both headphone types, as well as the means in each sound scene separately, are shown in Figure 3. Across all sound scenes, all means were significantly different with respect to each other, except the means of (noiseless) and (proposed).

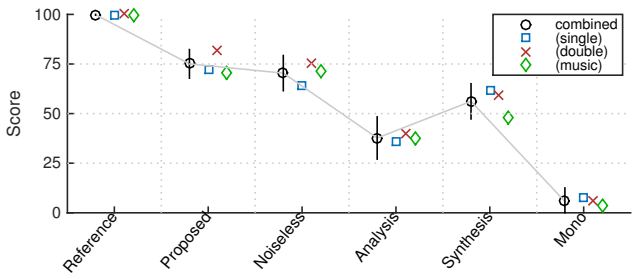

Fig. 3. Means and 95% confidence intervals of the overall results, and the means with each reference scene separately.

The result analysis shows that the proposed technique provides the perceptual quality over that obtainable by the linear means. The (analysis) mode suffers from noise amplification, and the (synthesis) mode suffers from the lack of directivity, which was readily observable in the purposefully wide sound scenes applied in the listening test. The (proposed) mode performed equally well to the (noiseless) mode. This was according to the design goal, since the (proposed) mode aims to provide the spatial quality of the (analysis) mode, however, without the negative effect of microphone self noise amplification. Many further factors could be studied for a more complete picture of the present state of quality of the proposed method, such as reducing the array size further, varying the microphone number and/or varying the amount of microphone noise. Nevertheless, it is expectable that the proposed parametric technique can be applied for benefit in any configuration due to its feature of combining the sound quality and spatial selectivity beyond that obtainable with linear means.

5. CONCLUSION

A perceptually motivated binaural rendering technique from microphone signals was proposed that achieves accurate spatial characteristics to the reproduced sound, however with high robustness against microphone self noise amplification. Beam patterns according to HRTFs are applied for parameter analysis. Signals from wider low-noise beam patterns are processed in time-frequency domain to have the spatial properties accordingly. Results of the listening test confirmed that the proposed method provided the same overall performance than a linear technique that was idealized in terms of being not subject to microphone noise. Thus, the proposed technique allows preserving high reproduction quality while reducing the microphone array size beyond the point in which the quality with linear techniques is compromized. The set of samples utilized in the listening test is provided in www.acoustics.hut.fi/projects/mbr/soundExamples/index.html.

6. REFERENCES

[1] A. Poletti and U. P. Svensson, "Beamforming synthesis of binaural responses from computer simulations of acoustic spaces," *The Journal of the Acoustical Society of America*, vol. 124, no. 301, 2008.

[2] Z. Li and R. Duraiswami, "Headphone-based reproduction of 3D auditory scene captured by spherical / hemispherical microphone arrays," *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 5, May 2006.

[3] R. Duraiswami, "Hemispherical microphone arrays for sound capture and beamforming," *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 106–109, October 2005.

[4] T. D. Abhayapala and D. B. Ward, "Theory and design of high order sound field microphones using spherical microphone array," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, May 2002.

[5] B. Rafaely, B. Weiss, and E. Bachmat, "Spatial aliasing in spherical microphone arrays," *IEEE Transactions on Signal Processing*, vol. 55, no. 3, pp. 1003–1010, 2007.

[6] B. Bernschütz, A. Vázquez, Giner, C. Pörschmann, and J. Arend, "Binaural reproduction of plane waves with reduced modal order," *Acta Acustica united with Acustica*, vol. 100, no. 5, pp. 972–983, 2014.

[7] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The CIPIC HRTF database," *IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics*, pp. 99–102, October 2001.

[8] J. Plogsties, P. Minnaar, S. K. Olesen, F. Christensen, and H. Møller, "Audibility of all-pass components in head-related transfer functions," in *Audio Engineering Society Convention 108*, February 2000.

[9] E. Zwicker, "Subdivision of the audible frequency range into critical bands (frequenzgruppen)," *Journal of the Acoustical Society of America*, vol. 33, no. 2, pp. 248–248, 1961.

[10] J. Herre, K. Kjörling, J. Breebaart, C. Faller, S. Disch, H. Purnhagen, J. Koppens, J. Hilpert, J. Rödén, W. Oomen, K. Linzmeier, and K. S. Chong, "MPEG Surround - the ISO/MPEG standard for efficient and compatible multichannel audio coding," *Journal of the Audio Engineering Society*, vol. 56, no. 11, pp. 932–955, 2008.

[11] J. Breebaart, S. van de Par, A. Kohlrausch, and E. Schuijers, "Parametric coding of stereo audio," *EURASIP Journal on Applied Signal Processing*, pp. 1305–1322, 2005.

[12] J. Herre, H. Purnhagen, J. Koppens, O. Hellmuth, J. Engdegård, J. Hilpert, L. Villemoes, L. Terentiv, C. Falch, A. Hölzer, M. L. Valero, B. Resch, H. Mundt, and H.-O. Oh, "MPEG Spatial Audio Object Coding - the ISO/MPEG standard for efficient coding of interactive audio scenes," *Journal of the Audio Engineering Society*, vol. 60, no. 9, pp. 655–673, 2012.

[13] P. Ekstrand, "Bandwidth extension of audio signals by spectral band replication," in *1st IEEE Benelux Workshop on Model based Processing and Coding of Audio*, November 2002.

[14] E. Vickers, "Frequency-domain implementation of time-varying FIR filters," in *Audio Engineering Society Convention 133*, October 2012.

[15] P.-A. Gauthier, C. Camier, Y. Pasco, A. Berry, E. Chambatte, R. Lapointe, and M.-A. Delalay, "Beamforming regularization matrix and inverse problems applied to sound field measurement and extrapolation using microphone array," *Journal of Sound and Vibration*, vol. 330, no. 24, pp. 5852–5877, August 2011.

[16] J. Vilkamo, T. Bäckström, and A. Kuntz, "Optimized covariance domain framework for time–frequency processing of spatial audio," *Journal of the Audio Engineering Society*, vol. 61, no. 6, pp. 403–411, 2013.

[17] *IBM Corp. Released 2013. IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp.*