
This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.

Author(s): Archontis Politis, Mikko-Ville Laitinen, Jukka Ahonen, Ville Pulkki
Title: Parametric Spatial Audio Processing of Spaced Microphone Array Recordings for Multichannel Reproduction
Year: 2015
Version: Final published version

Please cite the original version:

Archontis Politis, Mikko-Ville Laitinen, Jukka Ahonen, Ville Pulkki. Parametric Spatial Audio Processing of Spaced Microphone Array Recordings for Multichannel Reproduction. Journal of the Audio Engineering Society, 63, 4, 216-227, April 2015. DOI: 10.17743/jaes.2015.0015

Rights: © 2015 The Audio Engineering Society. Reprinted with permission.

This publication is included in the electronic version of the article dissertation:
Politis, Archontis. Microphone array processing for parametric spatial audio techniques.
Aalto University publication series DOCTORAL DISSERTATIONS, 195/2016.

All material supplied via Aaltodoc is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

Parametric Spatial Audio Processing of Spaced Microphone Array Recordings for Multichannel Reproduction*

ARCHONTIS POLITIS,¹ *AES Student Member*, MIKKO-VILLE LAITINEN,^{1†} *AES Associate Member*,
(archontis.politis@aalto.fi) (mikko-ville.laitinen@aalto.fi)

JUKKA AHONEN,² *AES Member*, AND VILLE PULKKI,¹ *AES Fellow*
(jukka.ahonen@akukon.fi) (ville.pulkki@aalto.fi)

¹*Aalto University, Department of Signal Processing and Acoustics, FI-02150 Espoo, Finland*

²*Akukon, FI-00380 Helsinki, Finland*

Recordings of music performances with spaced microphone array setups, when reproduced in a multichannel system, exhibit a good perceptual enveloping quality due to low inter-channel coherence in reverberant conditions, at the expense of stable localization of discrete sound events. In this work a parametric processing method is presented for spaced microphone array recordings using the principles of Directional Audio Coding (DirAC) based on knowledge of the array configuration and the directional response of the microphones. The method achieves equal or better quality to the standard high-quality version of DirAC, and it improves the common direct playback of such recordings by offering improved and stable localization cues and reduction of coloration issues at all frequencies.

0 INTRODUCTION

After the adoption of stereophony in the music production industry and its evolution to multichannel loudspeaker reproduction, such as early quadruphony or the later standardized surround setups, multichannel recording techniques have been an active topic of research. Setting aside individual recording of instruments with spot-microphones and their mixing by the sound engineer, capturing faithfully the overall sound scene requires carefully placed microphone arrays, which are intended for direct reproduction on the surround system.

There are two major categories of recording approaches for multichannel sound reproduction, coincident, and spaced techniques. Coincident techniques assume microphones with very small spacings in the range of a few centimeters that are “seen” by the wavelengths of interest as being in the same point in space. Coincident techniques assume that microphones are very closely placed, in the range of a few centimeters, and that they are “seen” by the wavelengths of interest as being in the same point in space.

A single impinging wavefront is then captured by all microphones with close-to-equal phase but with different am-

plitudes, depending on the orientation of the microphones and their directivity patterns. Examples of coincident techniques are the well-known mid-side stereophony, its multichannel extension of double mid-side [2], and the first-order B-format signals, which form the basis for the ambisonics method of sound capture and reproduction pioneered by Gerzon [3]. Coincident methods distribute the audio to the loudspeakers by means of matrixing the microphone channels, and they also permit some control of the reproduced stage-width and other directional characteristics of the perceived sound scene after the recording has been made. Furthermore, due to the fact that they encode primarily inter-channel level differences they can achieve consistent localization of discrete sound events at all directions around the listener. However, in terms of perception, the drawback of coincident recordings is the high correlation of the output channels. Even in the case of mostly diffuse sound, the reproduced channels are highly correlated for all frequencies resulting in a perceived lack of envelopment and depth, loss of immersion, and comb filter-like effects for small movements of the head.

In the second class of recording methods, that of spaced arrays, the microphones have considerable distances between them, so that an impinging wavefront is captured by the microphones with different time delays. As Theile presents [4], in the case of directional microphones a combination of inter-channel phase and level differences occur, which is determined by the direction of incidence, the geometry of the array, the directional patterns of the

*This work was originally presented in 134th AES Convention, Rome, Italy, 2013 May 4–7, in the form found in [1], and was awarded the Best Student Technical Paper award. It was revised for this publication in October 2014.

[†]Currently with Nokia Technologies, FI-02150 Espoo, Finland

microphones, and their orientation. Kassier argues [5] that, in one-to-one channel mapping between recordings and reproduction as is usually the case, these time and level differences manifest themselves as a more diffuse enveloping listening experience compared to coincident techniques, due to a reduced inter-aural coherence. However, Pulkki [6] shows that this is achieved at the expense of accurate localization of spatially discrete sound sources, which may sound either spread in space or with a fluctuating direction for most common arrays in use.

A recent parametric technique for multichannel audio, Directional Audio Coding (DirAC) [7], is based on analyzing B-format recordings and extracting psychoacoustically relevant directional information, which is subsequently used in the synthesis stage to generate the output channels in a way that, when reproduced, they are perceptually close to the original. From an audio-engineering perspective, DirAC aims to combine the best of both approaches, coincident and spaced recordings, through a separation between the directional and the non-directional/diffuse sounds. This separation enables accurate and stable localization along with good sense of envelopment and depth in reproduction.

The aim of this work is twofold. The first objective is to combine the best qualities of both recording classes by enhancing the spaced-microphone recordings in reproduction with stable and clear directional cues, without losing their open enveloping quality. This is achieved by a parametric analysis/synthesis stage based on DirAC. The second objective is to improve the parametric processing of DirAC by utilizing the spaced recordings. The suggested method improves and generalizes previous work presented by Laitinen et al. [8], which employed spaced omnidirectional microphones with DirAC, but was dependent on a simultaneous B-format recording for the directional analysis. Instead, a new formulation is proposed, intended for arrays of directional microphones for music recording, without additional B-format microphones. Furthermore, the proposed extraction of the directional information in the recording permits a unified common representation of both coincident and spaced recording formats and new possibilities for creative processing/mixing of the material [9].

The paper is organized as follows. In Sec. 1 a short description of spaced microphone arrays and the basics of DirAC are presented. Sec. 2 presents the analysis methods to obtain the directional parameters from the array recordings. In Sec. 3 the synthesis method based on the analyzed parameters is described for the reproduction of the recordings. Finally, Sec. 4 describes a practical implementation based on a simulated example array with realistic responses, and Sec. 5 presents formal listening test results for the same array setup.

1 BACKGROUND

1.1 Spaced Microphone Arrays

In the case of spaced array arrangements, the perceived sound scene depends completely on the adjustment of inter-microphone distances, pick-up patterns, and orientations of

the individual microphones. Since various objectives should be met throughout the design process, such as perceived stage width, clarity of imaging on the front, diffuse reproduction on the back, etc., there are many different arrangements in use. Examples of some established arrangements can be seen in Fig. 1, where all dimensions and orientations are based on the ones provided by Theile [4]. Kassier et al. [5] attempts a further categorization of the various designs in use, with a distinction between “main-microphone techniques” and “techniques with front-rear separation.” Main-microphone techniques are the ones where a single spaced array with front and surround microphones is used for capturing the sound scene, usually placed close to the music stage, approximately on the critical distance from the instruments.

Front-rear techniques use instead a main 3-channel array placed close to the stage, intended to capture mostly the direct sound for the front-channels (L-C-R) and a second multichannel array placed several meters away from the main one and adjusted to capture mostly reverberant sound. Then the total output in reproduction is a mix between the front and back arrays. Examples of front arrays are the popular Fukada tree [10], the Optimal Cardioid Triangle (OCT) variants [4], the INA-3 variants [4]. Examples of rear arrays are the Hamasaki square [11] and the IRT-Cross array [4]. The front arrays can also be augmented to full-surround main arrays by the addition of surround microphones, such as OCT-surround and INA-5 arrangements [4].

Various researchers have proposed psychoacoustic curves that make a link between time-delays and microphone directionality and perceived localization for various frontal angles of incidence, such as the early Franssen curves [12] for stereophonic and the ones given by Theile [4] and Williams [13] for multichannel front arrays. Recently, De Sena [14] provided an objective evaluation of recording and reproduction based on these principles, by observing the reproduced intensity field around the listening spot, and derived rules to optimize the microphone arrangements and directivities according to them.

1.2 Directional Audio Coding

DirAC is a parametric method for a perceptually motivated representation, transmission, and reproduction of spatial sound [7]. The model of DirAC assumes that with a time-frequency representation similar or finer to the resolution of the auditory system, it is adequate to encode and decode the local sound field with a reduced set of audio streams, compared to the output channels, and two parameters for each time-frequency tile. These are the direction of arrival (DOA) of incident sound energy and the diffuseness. DOA is assumed to relate to the directional cues of localization, while diffuseness relates to the sense of reverberation or sound source extent represented by interaural coherence. These parameters are used in the synthesis stage to recreate perceptually the captured sound scene. In traditional DirAC processing, the parameters are extracted in the time-frequency domain with an analysis based on the sound intensity and energy density at the recording position.

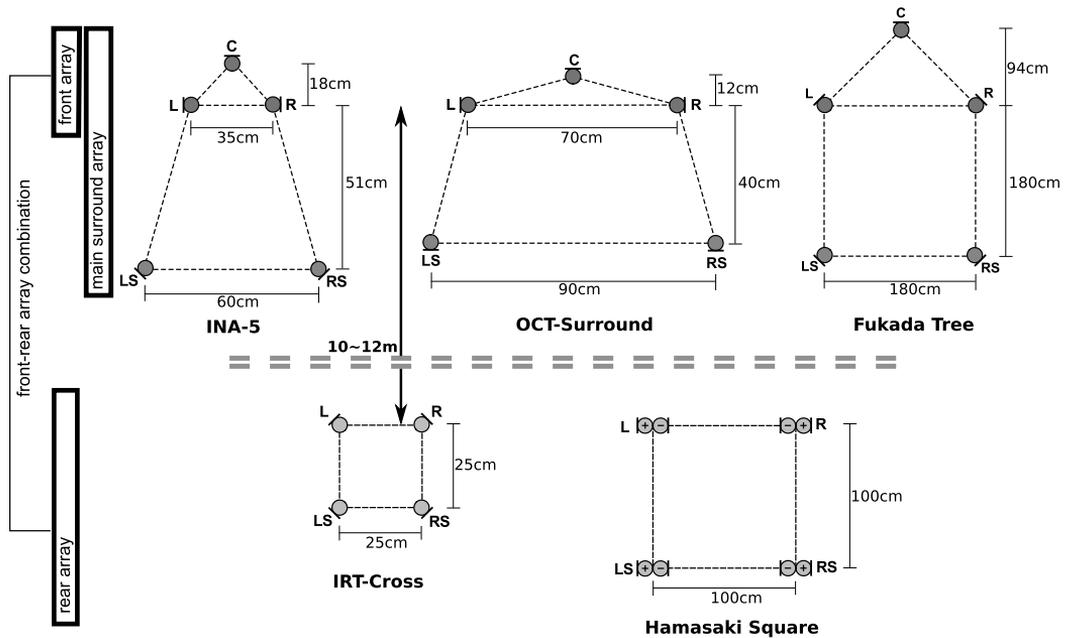


Fig. 1. Example spaced microphone arrays in use. Front-array arrangements indicate triplets of microphones for capturing the sound events from the main stage, which constitute main surround arrays when they are extended with surround microphones. Rear arrays refer to arrangements intended to capture diffuse sound.

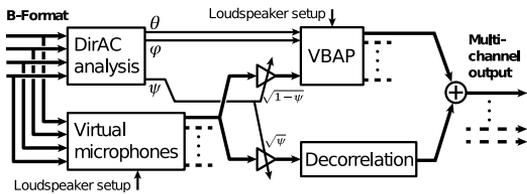


Fig. 2. Block diagram of DirAC processing for a single sub-band and a single loudspeaker channel. Dashed lines correspond to the rest of the loudspeaker channels. The time-frequency transform before and after the processing is omitted in the figure.

These physical variables are extracted from the B-format signals.

In the synthesis stage, a virtual-microphone signal is computed as a weighted sum of B-format signals for each loudspeaker direction. Commonly, the virtual microphones correspond to first-order directivity patterns (e.g., supercardioids). The virtual-microphone signals are then manipulated according to the metadata. Diffuseness and DOA are used to generate two separate streams, the non-diffuse stream corresponding to the directional part of the recording and the diffuse stream corresponding to the diffuse part, as in Fig. 2. Their respective mixing gain is determined by the diffuseness value. The directional part is then reproduced by means of the vector-base amplitude panning (VBAP) method, operating separately on each time-frequency bin or sub-band. The diffuse part is reproduced by means of

a separate decorrelation filter for each loudspeaker output. The final output is then the sum of the two streams for each loudspeaker.

1.3 Problem Statement

Listening tests by Vilkkamo et al. [15] show that DirAC achieves good perceived audio quality in most cases. However, recently it has been found by Laitinen et al. that there also exist problematic cases, such as multiple simultaneous talkers in acoustically dry spaces [16] and applause-type signals [17]. A major source of the artifacts is the effect of decorrelators. Decorrelation scrambles the phase spectrum, which can cause perceivable differences with certain signals [18–20], such as transients sounding smeared and speech perceived to be more reverberant and distant than in reality. Direct playback of spaced-microphone recordings does not suffer from these problems at mid-high frequencies since the signals are incoherent enough. However, Pulkki [6] showed that combined time-differences from all playback channels manifest themselves as localization blurring. In addition, the microphone signals are coherent at low frequencies, which causes timbral and spatial artifacts [7]. A combination of both methods is expected to (a) improve the problematic cases of DirAC, by using the natural incoherence of the array recordings when possible and hence minimizing decorrelation, (b) enhance localization aspects of direct playback by the parametric approach of DirAC, and (c) eliminate coloration issues of direct playback by appropriate decorrelation at low frequencies.

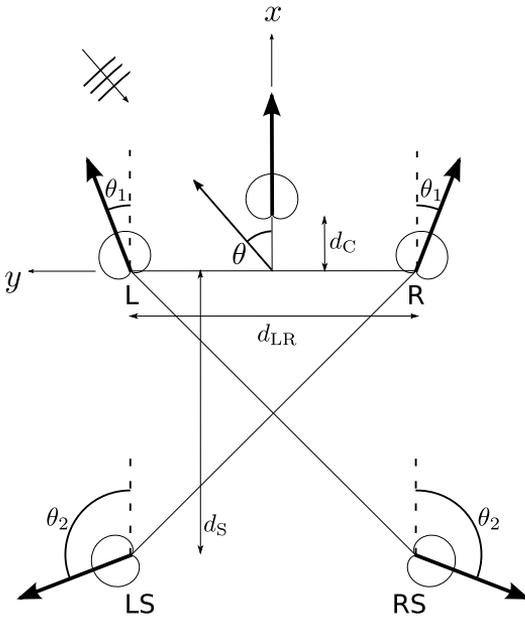


Fig. 3. Main surround array properties.

2 ANALYSIS AND PARAMETERIZATION OF SPACED ARRAY RECORDINGS

Previous work by Laitinen et al. [8] proposed use of a simultaneous B-format recording in the center of a spaced omnidirectional array for the directional analysis. An additional B-format microphone increases significantly the complexity and the cost of the array. Instead, it is preferred that all parameters are extracted directly from the spaced array recordings.

The DOA estimation in DirAC is based on direction of the intensity vector for each time-frequency block. Due to the large spacing, spaced arrangements such as the ones in Fig. 1 would allow estimation of intensity only at very low frequencies. Since our narrowband model assumes a mixture of a single plane wave and a diffuse field, a DOA estimator should be able to detect the direction of a single plane wave in the presence of diffuse sound after appropriate averaging. We propose a simple and efficient single-DOA estimator, with similar properties to the intensity one coming from the B-format analysis but based on the magnitude spectra of the microphone signals. This magnitude-based DOA estimation has been used previously by Politis et al. [21] for a tetrahedral array of cardioid microphones at frequencies above the aliasing limit of the array and a closely related estimation for a square array and power-spectrum differences has been presented by Ahonen et al. [22].

In the formulation the following assumptions are made. A general left-right symmetric array is assumed, which is always true in practice. The orientations of the microphones are parameterized according to the angles θ_1 for the stereo pair and the angles θ_2 for the surround pair, as seen in Fig. 3. Positive first-order patterns are assumed such as car-

dioids or subcardioids, with their pattern given by $D(\theta) = \beta + (1 - \beta)\cos\theta$, with $1 \geq \beta \geq 0.5$. All operations are defined in the time-frequency transformed domain, such as the short-time Fourier transform (STFT), for the time frame index n and frequency index k . For a plane wave of amplitude $a(k, n)$ incident from an angle θ , the signals captured by the array microphones are

$$\mathbf{s}(k, 4n) = \begin{bmatrix} s_L(k, n) \\ s_R(k, n) \\ s_C(k, n) \\ s_{LS}(k, n) \\ s_{RS}(k, n) \end{bmatrix} = a(k, n) \begin{bmatrix} e^{-j\varphi_L} D(\theta - \theta_1) \\ e^{-j\varphi_C} D(\theta) \\ e^{-j\varphi_R} D(\theta + \theta_1) \\ e^{-j\varphi_{LS}} D(\theta - \theta_2) \\ e^{-j\varphi_{RS}} D(\theta + \theta_2) \end{bmatrix} \quad (1)$$

with $\varphi_L, \varphi_C, \varphi_R, \varphi_{LS}, \varphi_{RS}$ being the respective phase differences due to the position of each microphone and DOA of the plane wave. We further define the vector of magnitude spectra for the microphone signals as

$$\mathbf{m}(k, n) = \begin{bmatrix} |s_L(k, n)| \\ |s_R(k, n)| \\ |s_C(k, n)| \\ |s_{LS}(k, n)| \\ |s_{RS}(k, n)| \end{bmatrix} = |a(k, n)| \begin{bmatrix} D(\theta - \theta_1) \\ D(\theta) \\ D(\theta + \theta_1) \\ D(\theta - \theta_2) \\ D(\theta + \theta_2) \end{bmatrix}. \quad (2)$$

Assuming that the time frame length is significantly larger than the temporal dimensions of the array, \mathbf{m} depends only on the magnitude of the plane wave and the gains of each microphone due to their directivity.

A DOA vector is defined

$$\mathbf{i}_{\text{doa}} = |a(k, n)| \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} \quad (3)$$

that is proportional to the magnitude of the plane-wave signal and points to the DOA θ . Two methods are presented for the estimation of this DOA vector; one for a main surround array (Sec. 2.1) and another using only the front triangle array (Sec. 2.2). When diffuse sound is also present, the direction of the DOA vector is going to fluctuate around the true DOA with a variance that depends on the power of the diffuse sound compared to the plane wave power. This fact is further exploited to obtain a diffuseness estimate.

At low frequencies, where the spacing between the microphones is significantly smaller than the wavelength, the microphones can be assumed coincident. It is then possible to obtain B-format signals from the first-order microphones and perform an intensity-based estimation as normally done in DirAC. This combination of two DOA estimators, the intensity-based one and the proposed magnitude-based one at different frequency ranges, has been used previously by the authors [21;22]. For simplicity in this work the magnitude-based estimator is used for all frequencies.

2.1 DOA Estimation—Main Surround Array Analysis

When a main surround array is used, the DOA vector can be estimated by using opposite pairs of the front and

surround microphones (e.g., the L-RS/R-LS pairs). The formulation is based on magnitude spectra differences of these opposing pairs that after some trigonometric manipulations result in dipole-like patterns scaled by the magnitude of the plane wave signal, similar to Eq. (3). More specifically,

$$\mathbf{i}_{\text{doa}}(k, n) = \mathbf{G}_{\text{msa}} \mathbf{m}(k, n) \quad (4)$$

where \mathbf{G}_{msa} is the matrix that subtracts and normalizes the spectra to result in the form of Eq. (3). With respect to the array parameters it is given by

$$\mathbf{G}_{\text{msa}} = \begin{bmatrix} g_1 & 0 & g_1 & -g_1 & -g_1 \\ -g_2 & 0 & g_2 & -g_2 & g_2 \end{bmatrix} \quad (5)$$

with

$$g_1 = \frac{1}{4(\beta - 1) \sin\left(\frac{\theta_1 + \theta_2}{2}\right) \sin\left(\frac{\theta_1 - \theta_2}{2}\right)}$$

$$g_2 = \frac{1}{4(\beta - 1) \sin\left(\frac{\theta_1 + \theta_2}{2}\right) \cos\left(\frac{\theta_1 - \theta_2}{2}\right)}. \quad (6)$$

It is obvious from Eq. (6) that the estimator requires the microphones to be directional and fails for omnidirectional ones with $\beta = 1$.

2.2 DOA Estimation—Front Array Analysis

For front-rear separated arrays an alternative formulation for the DOA estimation is given that uses only the front triangle of microphones. Hence, this formulation is suitable for both main surround and front-rear separated arrays.

Following a similar analysis as for the main surround array case, the DOA vector is computed from the signal magnitudes as

$$\mathbf{i}_{\text{doa}}(k, n) = \mathbf{G}_{\text{fa}} \mathbf{m}_{\text{fa}}(k, n) \quad (7)$$

where $\mathbf{m}_{\text{fa}} = [|\mathbf{s}_L|, |\mathbf{s}_C|, |\mathbf{s}_R|]^T$ is the magnitude spectra of the front array. The matrix \mathbf{G}_{fa} in this case is

$$\mathbf{G}_{\text{fa}} = \begin{bmatrix} -g_1 & 2g_1 & -g_1 \\ -g_2 & 0 & g_2 \end{bmatrix} \quad (8)$$

with

$$g_1 = \frac{1}{2(\beta - 1)(\cos \theta_1 - 1)}$$

$$g_2 = \frac{1}{2(\beta - 1) \sin \theta_1}. \quad (9)$$

Apart from the omnidirectional case, the DOA estimator using the front triangle array fails for orientations of the stereo pair parallel to the center microphone ($\theta_1 = 0$) since the y -axis dipole cannot be formulated, see Eq. (9). This limitation does not pose a problem since the angle of the stereo pair is never set to zero in practice.

2.3 Diffuseness Estimation

In the standard DirAC, diffuseness is estimated from the B-format signals [7]. Since computing the B-format is omitted with a spaced setup, or otherwise limited at very low frequencies, we use the alternative diffuseness estimator proposed by Ahonen et al. [23]. It is based on the temporal variation of the intensity vector for the case of direct-plus-diffuse sound. In this work the intensity vector is replaced

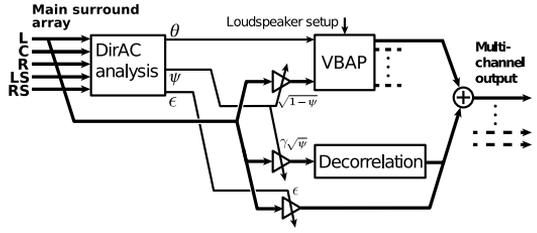


Fig. 4. Block diagram of DirAC processing for a main surround array. A single sub-band and a single loudspeaker channel is shown, as in Fig. 2.

by the DOA vector of Eq. (3) and the diffuseness is given by

$$\psi(k, n) = \sqrt{1 - \frac{|\mathbf{E}\{\mathbf{i}_{\text{doa}}(k, n)\}|}{\mathbf{E}\{|\mathbf{i}_{\text{doa}}(k, n)|\}}}, \quad (10)$$

where \mathbf{E} denotes statistical expectation. In practice the expectation is realized with time-averaging.

3 SYNTHESIS OF SPACED ARRAY RECORDINGS

The synthesis stage of the spaced-microphone processing needs to be adapted differently for a main surround array and for a front-rear combined array. The main emphasis in this work is on synthesis of main surround array recordings, which is more practical since only a single array needs to be set up in the performance space. In addition, the proposed method optimizes the rendering of the reverberant sound from the main array so that a rear array dedicated to just reverberant sound becomes redundant. However, in case a front-rear array is used, general guidelines are provided in Sec. 3.3 on how to process the recordings using a simplified version of the proposed synthesis.

The synthesis of main array recordings shares the same overall flow of standard DirAC with the virtual microphones replaced by the spaced microphones, see Fig. 4. However, significant modifications are introduced to both the rendering of the directional and the diffuse stream, in order to process the spaced-microphone signals in an optimal way.

3.1 Directional Sound

The part of the recording that is analyzed as directional, expressed as directional stream (DIRs), can be reproduced as in standard DirAC with virtual microphones, where the output signal of each loudspeaker is a combination of the virtual-microphone signals and VBAP gains. More specifically, the output signals can be given as

$$y_i(k, n) = \sqrt{1 - \psi(k, n)} g_i(k, n) x_i(k, n), \quad (11)$$

where $g_i(k, n) = g_i^{\text{VB}}(\theta(k, n))$ is the VBAP gain for the i^{th} loudspeaker and DOA $\theta(k, n)$, and $x_i(k, n)$ is the signal of the i^{th} virtual or spaced microphone.

This formulation ignores the effect of the patterns of the microphones and their orientation, virtual or not. In the case of virtual microphones the effect is partially mitigated by the frequency-independent behavior of the patterns up to the operating frequency of the B-format microphone, and the fact that the microphones are always oriented toward the loudspeakers, which permits an average statistical correction [15]. This correction cannot correct deviations in loudness of a single moving source, as it moves from the direction of a loudspeaker to between a pair or a triplet of them. In the case of spaced microphones this effect is more pronounced due to the combination of both level differences and time differences occurring between the reproduced channels, which can result in both power loss and directional bias at the listening spot for certain directions. The effect of the time-differences can be treated appropriately in the design of the spaced array, see Sec. 4.3.

Based on knowledge of the microphone responses and the estimated DOA from the analysis stage, we propose an exact solution to achieve constant power with direction. Ideally, for a single frequency bin k the sum of the powers of the combined gains of all speakers and microphones for each direction should be unity, $\sum g_i^2(\theta) = 1$. To achieve this normalization, we introduce a frequency-dependent compensation $c(\theta, k)$ along with the gains so that the sum becomes

$$\sum_{i=1}^M [c(\theta, k) g_i^{\text{VB}}(\theta) m_i(\theta, k)]^2 = 1 \quad (12)$$

with m_i the microphone gain due to the i^{th} frequency-dependent pattern. Then the compensation gain is simply given by

$$c(\theta, k) = \frac{1}{\sqrt{\sum_{i=1}^M [g_i^{\text{VB}}(\theta) m_i(\theta, k)]^2}}. \quad (13)$$

Using this correction, the properly normalized output for a single loudspeaker becomes

$$y_i(k, n) = \sqrt{1 - \psi(k, n)} \hat{g}_i(k, n) x_i(k, n) \quad (14)$$

with $\hat{g}_i(k, n) = c(\theta(k, n), k) g_i^{\text{VB}}(\theta(k, n))$.

3.2 Diffuse Sound

As has already been mentioned, a clear advantage of spaced-microphone input for the parametric processing is the fact that inter-channel coherence is much lower, especially at mid-high frequencies, than the respective virtual microphones in the B-format processing. This natural incoherence can be exploited to reduce the amount of decorrelation and its effects on the quality of reproduction. To achieve that, the diffuse stream is further split into two streams: a decorrelated and a non-decorrelated one. The decorrelated stream (DIFFs) is similar to the normal diffuse stream of the B-format processing, however its portion of the total energy of the signal is smaller and this portion is frequency-dependent. The non-decorrelated stream (NDECs) is routed with a one-to-one mapping

relation from the microphones to the loudspeakers. The mixing ratio between the two is determined by the coherence between all pairs of microphones in a diffuse field.

Assuming that the coherence curves are known between all pairs of microphones, different methods can be followed for the decomposition of the DIFFs into decorrelated and non-decorrelated. Vilkamo et al. [24] presents a closely related method that performs this decomposition implicitly through an estimation of the covariance matrix of the input channels. However, in that method all input channels may contribute to a single output channel and this mixing process is adaptive. The approach presented herein is based on an average coherence curve γ_{av} derived from all pairs of microphones in the array. Then the power ratio between DIFFs and NDECs for each output channel is based on this static average curve.

In the standard DirAC processing, the output signals of the DIFFs are given by

$$y_i(k, n) = \sqrt{\psi(k, n)} \sqrt{Q_{\text{vm}} \alpha_i} x_i(k, n) h_i^{\text{dec}}(k) \quad (15)$$

where α_i are per-loudspeaker weights that distribute evenly the energy of the diffuse field depending on the regularity of the speaker setup. They also result in a unit power reproduction for a unit power diffuse-field recording and they hold the relation $\sum_i \alpha_i^2 = 1$. The h_i^{dec} is an independent decorrelating filter per speaker output. The Q_{vm} is the directivity factor of the virtual microphones, employed in order to counteract the power reduction of the diffuse field captured by a directional microphone.

In the spaced array formulation, the DIFFs output is similar to Eq. (15). However, in contrast to the frequency-independent virtual microphones, in this formulation the frequency-dependent directivity factor of the array is taken into account, and Q_{vm} is replaced with $Q_{\text{sm}}(k)$ for the spaced microphones. Finally, at a single frequency bin k , the gains for the three streams are

$$\begin{aligned} G_i^{\text{dir}} &= \sqrt{1 - \psi} g_i^{\text{VB}}(\theta) m_i(\theta) c(\theta) \\ G_i^{\text{diff}} &= \gamma_{\text{av}} \sqrt{\psi} Q_{\text{sm}} \alpha_i \\ G_i^{\text{ndec}} &= \epsilon \alpha_i \end{aligned} \quad (16)$$

where the G_i^{dir} , G_i^{diff} , G_i^{ndec} are the gains for the DIRs, the DIFFs, and the NDECs, respectively. The quantity ϵ defines the amount of the NDECs and is determined below. Since the DIFFs outputs are decorrelated versions of the input signals, their energy sum incoherently with the DIRs outputs. However, the NDECs is completely coherent with the DIRs and thus, in order to preserve the energy of the input channels, the following condition has to be achieved

$$\sum_i (G_i^{\text{dir}} + G_i^{\text{ndec}})^2 + \sum_i (G_i^{\text{diff}})^2 = 1 \quad (17)$$

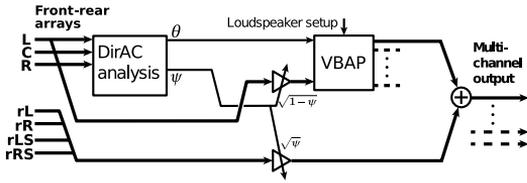


Fig. 5. Block diagram of DirAC processing for front-rear combined arrays. A single sub-band and a single loudspeaker channel is shown, as in Fig. 2.

By using Eqs. (16) and (17) and the relations $\sum_i (G_i^{\text{dir}})^2 = 1 - \psi$, $\sum_i (G_i^{\text{diff}})^2 = \gamma_{\text{av}}^2 \psi$ and $\sum_i (G_i^{\text{ndec}})^2 = \epsilon^2$, we arrive at the quadratic equation

$$\epsilon^2 + \left[2\sqrt{1-\psi} c(\theta) \sum_i (g_i^{\text{VB}}(\theta) m_i(\theta) \alpha_i) \right] \epsilon + \psi(\gamma_{\text{av}}^2 - 1) = 0 \quad (18)$$

which has a single valid solution that fulfills $\epsilon \geq 0$

$$\epsilon = \sqrt{A^2 - B} - A \quad (19)$$

where

$$A = 2\sqrt{1-\psi} c(\theta) \sum_i (g_i^{\text{VB}}(\theta) m_i(\theta) \alpha_i) \quad (20)$$

$$B = 4\psi(1 - \gamma_{\text{av}}^2)$$

It is obvious from Eq. (20) that the gain of the NDECs depends both on the estimated diffuseness and DOA.

3.3 Front-Rear Separated Arrays

Fig. 5 shows the proposed scheme for a front-rear separated array. In this case, only the rear array is used for the diffuse stream. Since the rear arrangements are positioned far enough from the stage, the reverberant sound dominates. It can then be assumed that the signals are completely incoherent with the directional stream, which originates from the front array close to the stage. Hence, decorrelation is not used in this case. Furthermore, the amplitude panning gains for the directional stream are limited in the range of the front-center speakers, since it is not possible to achieve spatial imaging on the side with panning due to low coherence between the L/R and the LS/RS channels. This is achieved simply by limiting the DOA parameter to the range between the L-R speakers even when the analyzed directions fall outside of it. A delay control based on the distance between the front and rear array can be also used on the front triangle signals to adjust the time-alignment between the directional stream and the diffuse stream.

4 PRACTICAL IMPLEMENTATION AND SIMULATIONS

In this section a practical implementation of the suggested spaced-microphone processing is presented along with simulation results. The method is demonstrated for the standard multichannel ITU 5.1 loudspeaker layout. The

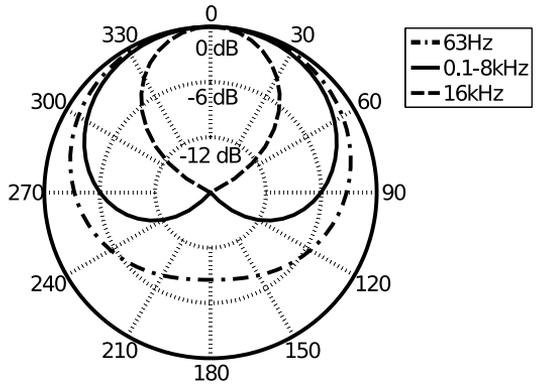


Fig. 6. Directivity patterns of the simulated cardioid microphone. The order of the cardioid rises linearly from $N=1$ at 8 kHz to $N=6$ at 16 kHz.

section starts by discussing the directivity and the coherence properties of microphones. Based on them, a suitable microphone array for spaced-DirAC processing is suggested. The performance of the suggested solution is validated with simulation results. Furthermore, the results of formal listening tests are presented in Sec. 5.

4.1 Simulated Microphone Capsules

As discussed in Sec. 2, cardioid patterns are a suitable choice for the directional analysis. The simulations presented below are based on realistic frequency-dependent cardioid microphones, where the directional patterns exhibit a cardioid response at mid frequencies (0.1 – 8 kHz), linearly-rising order cardioid response $D(\theta, N) = (0.5 + 0.5\cos\theta)^N$ at high frequencies (above 8 kHz), and at the low-end (below 0.1 kHz) they deviate toward omnidirectional with a linearly varying first-order response, see Fig. 6. This simple model was found to be close to the real response of a common cardioid used for music recording (AKG C414 XLS), measured in an anechoic chamber. The analytical model of the capsule directivities gives directly the knowledge of the pattern responses for a specific DOA and the directivity factor with frequency, which are used in the processing steps described previously.

4.2 Coherence between the Microphones

For diffuse sound fields, a well-known closed-form solution exists for the coherence between two omnidirectional microphones at a certain distance. For the three-dimensional case the coherence is given by [25]

$$\gamma_p(k, r) = \frac{\sin(kr)}{kr} \quad (21)$$

where k is the wavenumber, and r is the distance between the microphones. The signals are highly coherent at low frequencies and the coherence decays rapidly as frequency increases. The larger the spacing the lower the frequency at which the coherence starts to decay.

There exist analytical relations for the coherence between a pair of first-order microphones with arbitrary spacing and orientations and for higher-order differential microphones for collinear orientations [26]. In the case of complex but known frequency-dependent patterns at arbitrary orientations, such as the assumptions in this paper are, the coherence can be found numerically by simulation of the microphone outputs in an approximation of a diffuse field.

In general, directional microphones offer reduced coherence compared to the omnidirectional case and apart from the distance the coherence curve depends on the patterns and their orientation. However, the general trend of the curve is similar to the omnidirectional case. In [8], it was noticed that below ~ 300 Hz decorrelation does not cause perceivable differences. Thus, a pair of microphones optimized for the proposed parametric processing should have a spacing that results in low coherence above 300 Hz in order to avoid decorrelation at this region. For omnidirectional sensors that distance is ~ 40 cm.

4.3 Microphone Array Design

The parametric method presented in this study is suitable for various standard spaced arrays. However, it is also possible to tune the array arrangement in a way optimal for the parametric processing. The simulations and the rendered samples for the listening test were generated with this approach in mind. The design is based on a simple array where the microphones are oriented toward the loudspeakers of a 5.1 system but slightly modified to better fulfill requirements of the analysis and synthesis stages.

From the analysis side, the requirements are minimal distances between the microphones. From the synthesis side, the requirements are (a) average distances between the microphones should be at least ~ 40 cm and (b) time-differences between adjacent microphones should be minimized. Obviously, the requirements are contradicting and thus the optimal solution is a compromise. A suggested design can be seen Fig. 7. The spacing between the microphones is a compromise enabling enough incoherence without disturbing the analysis too much. Usually, the main part of the directional content in a musical recording is in the front. Thus, the front microphones are optimized from the synthesis point of view in order to avoid temporal and level differences that might affect the perceived directions.

4.4 Synthesis

The adjustment of the amplitude panning gains to take into account the directionality of the microphones has been presented in Sec. 3.1. In order to demonstrate its effect we simulate a single plane wave carrying a broadband white noise signal, coming from a specific direction. The signal is encoded for the simulated array of Fig. 7 and it is further decoded to a 5.0 loudspeaker setup at 0° , $\pm 30^\circ$, $\pm 110^\circ$. The noise signal has an average RMS value of one and ideally, the RMS value of the outputs should be equal to that value regardless of the direction of arrival, maintaining the same loudness as the captured sound. The output signals are computed for 36 directions, from 0° to 180° every 5° ,

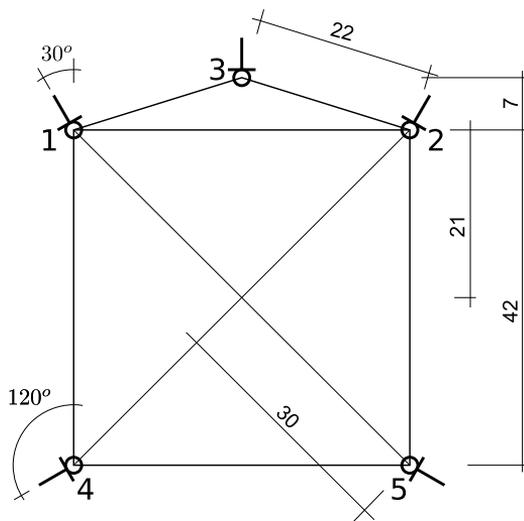


Fig. 7. The suggested microphone array. The look directions and the distances between the microphones are marked in the figure.

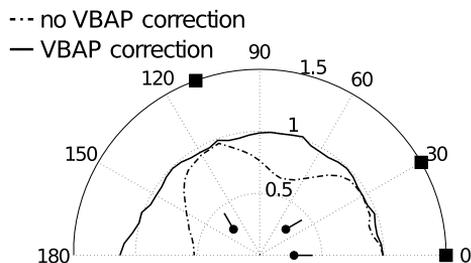


Fig. 8. RMS output on reproduction for a captured unit-power plane-wave, *dashed* line: no directional correction of VBAP gains, *solid* line: with directional correction of VBAP gains.

and their RMS value is plotted in Fig. 8. The loss of energy without the gain adjustment can be seen for directions on the sides and back.

Regarding the diffuse stream reproduction, the information of coherence between the microphone signals for all microphone pair combinations in a diffuse stream is needed in order to determine the average coherence curve for the method presented in Sec. 3.2. For N number of microphones there are $(N^2 - N)/2$ independent coherence functions. In our case, due to the symmetry of the specific array, only five coherence functions are needed to be known.

For the complex frequency-dependent patterns, such as the ones for the simulated array, the coherence is computed numerically. This is done by simulating the diffuse field signals for the specific microphones and by computing their power and cross-spectral densities, using, for example, Welch's periodogram method. The results can be seen in Fig. 9, where the simulated absolute coherences of the frequency-dependent patterns are plotted. Even though the coherence should decay to zero at higher frequencies, we

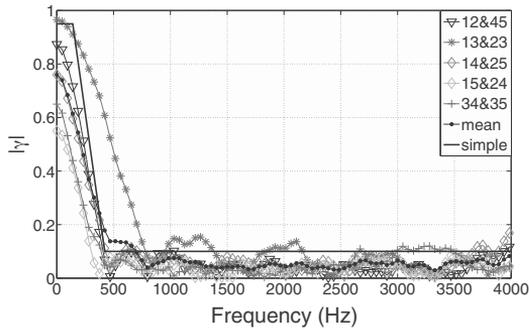


Fig. 9. Magnitude coherence between all microphones of the array of Fig. 7 for the frequency-dependent patterns of Fig. 6. The solid curves represent the possible coherences between all pairs of the array. The *mean* dashed curve denotes the average of these coherences, while the *simple* dashed curve is the simplified parametric curve used in the DirAC processing.

can see that there is an increasing estimation error due to limited frequency resolution and finite averaging of the periodograms. However, the plots are suitable to give an idea of the overall behavior of the coherence between the microphones. Along with the actual curves, their mean curve and a simplified parametric curve, which has been used in the spaced DirAC processing, are also plotted.

5 LISTENING TESTS

Formal listening tests were carried out in order to assess the perceived quality of reproduction using the method suggested in this paper. The aim is to study how well a sound scene can be captured and reproduced using the common 5.1 loudspeaker layout [27]. In addition, the suggested spaced-microphone DirAC reproduction is compared to normal DirAC reproduction and to spaced-microphone reproduction without any processing.

5.1 Listening Test Procedure

The listening test was performed in a listening room fulfilling the ITU-R BS.1116 recommendation [27]. Eighteen listeners participated in the test, excluding the authors of this paper. The loudspeaker setup included nine Genelec 8260A loudspeakers, which were positioned at 0° , $\pm 30^\circ$, $\pm 70^\circ$, $\pm 110^\circ$, and $\pm 160^\circ$ of azimuth at the elevation of 0° , and four Genelec 8240A loudspeakers, which were positioned at $\pm 45^\circ$ and $\pm 135^\circ$ of azimuth at the elevation of 45° . The listening position was in the sweet spot.

Three audio samples were used as the program material. Based on previous studies, the reproduction quality of DirAC processing is known to be good in most of the cases. Thus, challenging cases for DirAC processing, discussed in Sec. 1.3, were selected for this listening test. In the first sample, three simultaneous talkers were simulated inside a small room. The different talkers were positioned at -30° , 0° , and 110° . The second sample was otherwise identical, but the simulated room was large. The third sample included

Table 1. Properties of the simulated rooms.

Sample	Dimensions	Distance	T_{60}
<i>Speech, small</i>	8×6×3 m	2 m	0.8 s
<i>Speech, large</i>	50×20×18 m	8 m	2.0 s
<i>Music</i>	20×9.5×5 m	5 m	1.1 s

music simulated in a medium-size room. There were three instrument sections in the song, strings, percussion, and bass, positioned to -30° , 0° , and 30° , respectively. The simulated room responses were created with the image-source method for rectangular rooms. The properties of the simulated rooms used in the test samples can be seen in Table 1. The length of the samples was 16 seconds.

A reference case was created using the room simulator. Based on the simulator output, "ideal" microphone patterns were created for each loudspeaker in the reproduction layout. The ideal patterns corresponded to ideal non-overlapping sectors around each loudspeaker, covering the whole sphere. This was achieved by quantizing the direct sound and the reflections to the closest loudspeakers. The result is a 13-channel reference. The spaced-microphone and the B-format signals were simulated according to it. The layout of the spaced microphone array can be seen in Fig. 7.

There were five reproduction methods to be evaluated in the test. One of the test methods was a hidden reference (*ref*) that was identical to the 13-channel reference. A mono downmix of the loudspeaker channels of the reference (*mono*) acted as a low-quality anchor. In case of the other three methods, the reproduction was performed using a five-channel layout (0° , $\pm 30^\circ$, $\pm 110^\circ$). Using the spaced-microphone signals, the suggested spaced-microphone DirAC processing was applied in *spaced-DirAC*. Normal DirAC processing was applied to the B-format signals in *normDirAC*. The spaced-microphone signals were reproduced without any processing in *space-dRaw*.

The task of the subjects was to compare the overall quality of the reproduction methods to the reference. The question was: "Grade perceived impairment." The score was given using a continuous scale from 1 to 5 with a step size of 0.1. The following anchor words were used for describing the perceived impairment: 5 = *Imperceptible*, 4 = *Perceptible, but not annoying*, 3 = *Slightly annoying*, 2 = *Annoying*, and 1 = *Very annoying*. The subjects used a graphical user interface for selecting which sample to play and to write in the score. They listened to one sample at a time and graded different methods in a multiple-stimulus test. They were able to change between different methods freely and to seek desired excerpts in the samples. The order of the methods was randomized for each test case and the subjects did not know which method was used in a sample they were listening to. All samples were graded three times, and the order of the cases was randomized.

The listening test was conducted in two parts. The first part was a training session where the subjects did a short version of the actual listening test. This way the subjects

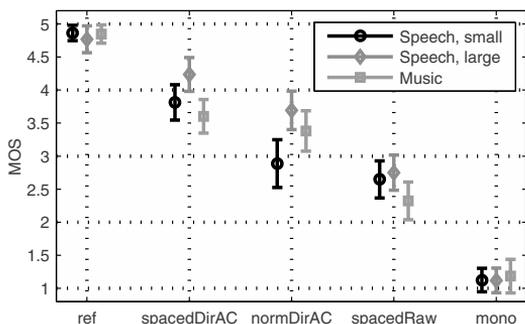


Fig. 10. Perceived impairment in the listening test. Means and 95% confidence intervals are shown.

became thoroughly familiar with the artifacts under study. The duration of the training sessions was 10 minutes on average, and the duration of the actual listening test was 30 minutes on average.

5.2 Results

Repeated-measures analysis of variance (RM-ANOVA) was used to obtain a statistical analysis of the results. The univariate approach was taken and the correction for the violation of sphericity was applied when required. The repetitions of the same combination were pre-averaged before the analysis.

Two-way RM-ANOVA was applied to the results. The within-subjects factors were *program material* and *method*. Mauchly's test of sphericity revealed that two factors violated the assumption of sphericity: *method* $\chi^2(9) = 18.33$, $p < 0.05$, $\epsilon < 0.75$; *sample-method* $\chi^2(9) = 61.35$, $p < 0.05$, $\epsilon < 0.75$. The correction method was Greenhouse-Geisser since $\epsilon < 0.75$. RM-ANOVA revealed the following significant factors: main effects *program material*, $F(2, 34) = 18.73$, $p < 0.05$, and *method*, $F(2.62, 44.49) = 309.85$, $p < 0.05$; and interaction *program material-method*, $F(4.62, 78.47) = 16.28$, $p < 0.05$. Fig. 10 shows mean opinion score (MOS) plots with 95% confidence intervals for the interaction.

The results show that *spacedDirAC* enables good quality with the MOS being close to 4 = *Perceptible, but not annoying* in all cases. In addition, quality is better or equally good as the traditional DirAC processing in all cases. It can also be seen that the suggested approach provides better quality than the direct playback of the spaced-microphone signals. It should be noted that the quality of the 5-channel reproduction methods was compared to a 13-channel reference.

6 CONCLUSIONS

A processing scheme for spaced-microphone recordings has been presented in this work that, according to the listening tests, achieves high-quality reproduction compared to both the unprocessed recordings and the standard DirAC method on which it is based and extends. By knowing the ar-

ray geometry and microphone characteristics, a directional-diffuse decomposition of the recordings is realized, which is subsequently used for the synthesis of the output channels. The processing scheme achieves improved localization at reproduction by combining the microphone directivities with amplitude-panning and retains pleasant enveloping qualities of spaced recordings by decorrelating the output signals only at the frequency range that is required. In addition, the proposed method extends the coincident parametric processing of DirAC to spaced-microphone input, which is an additional step towards a unification of recording multichannel formats into a common parametric representation.

REFERENCES

- [1] A. Politis, M.-V. Laitinen, J. Ahonen, and V. Pulkki, "Parametric Spatial Audio Coding for Spaced Microphone Array Recordings," presented at the *134th Convention of the Audio Engineering Society* (2013 May), convention paper 8905.
- [2] F. Rumsey, *Spatial Audio* (Focal Press, Oxford, 2001).
- [3] M. A. Gerzon, "Ambisonic in Multichannel Broadcasting and Video," *J. Audio Eng. Soc.*, vol. 33, pp. 859–871 (1985 Nov.).
- [4] G. Theile, "Multichannel Natural Recording Based on Psychoacoustic Principles," presented at the *106th Convention of the Audio Engineering Society* (2000 May), convention paper 4976.
- [5] R. Kassier, H.-K. Lee, and F. Rumsey, "An Informal Comparison between Surround-Sound Microphone Techniques," presented at the *118th Convention of the Audio Engineering Society* (2005 May), convention paper 6429.
- [6] V. Pulkki, "Microphone Techniques and Directional Quality of Sound Reproduction," presented at the *112th Convention of the Audio Engineering Society* (2002 Apr.), convention paper 5500.
- [7] V. Pulkki, "Spatial Sound Reproduction with Directional Audio Coding," *J. Audio Eng. Soc.*, vol. 55, pp. 503–516 (2007 Jun.).
- [8] M.-V. Laitinen, F. Kuech, and V. Pulkki, "Using Spaced Microphones with Directional Audio Coding," presented at the *130th Convention of the Audio Engineering Society* (2011 May), convention paper 8433.
- [9] A. Politis, T. Pihlajamäki, and V. Pulkki, "Parametric spatial audio effects," *15th Int. Conference on Digital Audio Effects (DAFx-12)*, York, UK (2012).
- [10] A. Fukada, "A Challenge in Multichannel Music Recording (Seminar Presentation)," presented at the *AES 19th Int. Conference: Surround Sound: Techniques, Technology, Perception* (2001 Jun.).
- [11] K. Hamasaki and K. Hiyama, "Reproducing Spatial Impression with Multichannel Audio," *AES 24th Int. Conference: Multichannel Audio, The New Reality* (2003 Jun.).
- [12] N. V. Franssen, "Stereophony," Technical Report, Philips Technical Library, Eindhoven, The Netherlands (1962).

- [13] M. Williams and G. Le Dû, "Microphone Array Analysis for Multichannel Sound Recording," presented at the *107th Convention of the Audio Engineering Society* (1999 Sep.), convention paper 4997.
- [14] E. De Sena, H. Hacihabiboglu, and Z. Cvetkovic, "Analysis and Design of Multichannel Systems for Perceptual Sound Field Reconstruction," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 8, pp. 1653–1665 (2013).
- [15] J. Vilkkamo, T. Lokki, and V. Pulkki, "Directional Audio Coding: Virtual Microphone Based Synthesis and Subjective Evaluation," *J. Audio Eng. Soc.*, vol. 57, pp. 709–724 (2009 Sep.).
- [16] M.-V. Laitinen and V. Pulkki, "Utilizing Instantaneous Direct-to-Reverberant Ratio in Parametric Spatial Audio Coding," presented at the *133rd Convention of the Audio Engineering Society* (2012 Oct.), convention paper 8804.
- [17] M.-V. Laitinen, F. Kuech, S. Disch, and V. Pulkki, "Reproducing Applause-Type Signals with Directional Audio Coding," *J. Audio Eng. Soc.*, vol. 59, pp. 29–43 (2011 Jan./Feb.).
- [18] R. D. Patterson, "A Pulse Ribbon Model of Monaural Phase Perception," *J. Acoust. Soc. Am.*, vol. 82, no. 5, pp. 1560–1586 (1987).
- [19] B. C. Moore, "Interference Effects and Phase Sensitivity in Hearing," *Philosophical Transactions of the Royal Society of London.*, vol. 360, 1794, pp. 833–858 (2002).
- [20] M.-V. Laitinen, S. Disch, and V. Pulkki, "Sensitivity of Human Hearing to Changes in Phase Spectrum," *J. Audio Eng. Soc.*, vol. 61, pp. 860–877 (2013 Nov.).
- [21] A. Politis and V. Pulkki, "Broadband Analysis and Synthesis for Directional Audio Coding Using A-Format Input Signals," presented at the *131st Convention of the Audio Engineering Society* (2010 Oct.), convention paper 8525.
- [22] J. Ahonen, G. D. Galdo, F. Kuech, and V. Pulkki, "Directional Analysis with Microphone Array Mounted on Rigid Cylinder for Directional Audio Coding," *J. Audio Eng. Soc.*, vol. 60, pp. 311–324 (2012 May).
- [23] J. Ahonen and V. Pulkki, "Diffuseness Estimation Using Temporal Variation of Intensity Vectors," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY, USA (2009).
- [24] J. Vilkkamo, T. Bäckström, and A. Kuntz, "Optimized Covariance Domain Framework for Time-Frequency Processing of Spatial Audio," *J. Audio Eng. Soc.*, vol. 61, pp. 403–411 (2013 Jun.).
- [25] F. Jacobsen, "The Diffuse Sound Field," Technical Report 27, Technical University of Denmark (1979).
- [26] W. G. Elko, "Spatial Coherence Functions for Differential Microphones in Isotropic Noise Fields," in M. Brandstein and D. B. Ward, eds., *Microphone Arrays*, chapter 3, pp. 61–85 (Springer-Verlag, Berlin, Heidelberg, New York, 2004).
- [27] ITU-R BS.1116, "Methods for the Subjective Assessment of Small Impairments in Audio Systems Including Multichannel Sound Systems" (1997).

THE AUTHORS



Archontis Politis



Mikko-Ville Laitinen



Jukka Ahonen



Ville Pulkki

Archontis Politis obtained his M.Eng. degree in civil engineering at Aristotle's University of Thessaloniki, Greece, and his M.Sc. degree in sound and vibration studies at ISVR, University of Southampton, UK, in 2006 and 2008 respectively. He has also studied piano performance and holds a degree in classical music studies from Alexandreion Conservatory, Thessaloniki, Greece. From 2008 to 2010 he worked as a graduate acoustic consultant at Arup Acoustics, Glasgow, UK, and as a researcher in a joint collaboration between Arup Acoustics and the Glasgow School of Arts on interactive auralization of architectural spaces using 3D sound techniques. Currently, he is pursuing a doctoral degree in the field of parametric spatial sound recording and reproduction in Aalto University, Espoo, Finland.

Mikko-Ville Laitinen received an M.Sc. (Tech.) degree, from the Helsinki University of Technology, Finland, in 2008, and a D.Sc. (Tech.) degree, from the Aalto University, Finland, in 2014. In 2006, he was working for the loudspeaker company Genelec, Finland. From 2007 to 2014 he was working as a researcher at the Aalto University, where his research topics included spatial audio, multichannel reproduction, psychoacoustics, and audio codecs. Furthermore, he was a visiting researcher at Fraunhofer IIS, Germany, in 2012. Currently he is working as a principal engineer at Nokia Technologies, Finland.

Jukka Ahonen received his D.Sc (Tech) degree from Aalto University, Espoo, Finland, in 2013. He has also graduated in music studies from the Helsinki Pop & Jazz Con-

servatory, Finland, in 2004. From 2006–2011 he worked as a researcher and Ph.D. candidate at Aalto University. In 2009 he was a visiting researcher at Fraunhofer IIS, Erlangen, Germany. His research interests include parametric spatial sound processing, spatial sound microphone techniques, and room acoustics. At present, he works as an acoustician and group manager at Akukon Ltd., a consulting company for technical acoustics, noise control, audio-visual design, and theater technics.

Ville Pulkki has been working in the field of audio from 1995. In his Ph.D. thesis (2001) he developed a method to position virtual sources for 3D loudspeaker set-ups after researching the method using psychoacoustic listening tests and binaural computational models of human hearing. Later he worked on the reproduction of recorded spatial sound scenarios, on the measurement of head-related acoustics, and on the measurement of room acoustics with laser-induced pressure pulses. Currently he holds a tenure-track assistant professor position in Aalto University and runs a research group with 14 researchers. He is a fellow of the Audio Engineering Society (AES) and has received the AES Publication Award. He has also received the Samuel L. Warner memorial medal from the Society of Motion Picture and Television Engineers (SMPTE). He has a background in music, having received teaching from the Sibelius Academy in singing and audio engineering alongside instruction in various musical instruments. He has also composed and arranged music for many different ensembles. He enjoys being with his family, renovating his summerhouse, and dancing to hip hop.