Audio Engineering Society

Convention Paper 7257

Presented at the 123rd Convention
2007 October 5–8 New York, NY

The papers at this Convention have been selected on the basis of a submitted abstract and extended precis that have been peer reviewed by at least two qualified anonymous reviewers. This convention paper has been reproduced from the author's advance manuscript, without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. Additional papers may be obtained by sending request and remittance to Audio Engineering Society, 60 East 42nd Street, New York, New York 10165-2520, USA; also see www.aes.org. All rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Energetic Sound Field Analysis of Stereo and Multichannel Loudspeaker Reproduction

Juha Merimaa[1]

[1] *Creative Advanced Technology Center, 1500 Green Hills Road, Scotts Valley, CA 95066, USA*

Correspondence should be addressed to Juha Merimaa (juham@atc.creative.com)

ABSTRACT

Energetic sound field analysis has been previously applied to encoding the spatial properties of multichannel signals. This paper contributes to the understanding of how stereo or multichannel loudspeaker signals transform into energetic sound field quantities. Expressions for the active intensity, energy density, and energetic diffuseness estimate are derived as a function of signal magnitudes, cross-correlations, and loudspeaker directions. It is shown that the active intensity vector can be expressed in terms of the Gerzon velocity and energy vectors, and its direction can be related to the tangent law of amplitude panning. Furthermore, several cases are identified where the energetic analysis data may not adequately represent the spatial properties of the original signals.

1. INTRODUCTION

Energetic sound field analysis has recently found applications in audio technology as a method for capturing the spatial properties of room responses [1, 2], surround recording and reproduction [3, 4, 5], or encoding of multichannel sound [4]. One of the advantages of such an approach is that, once encoded, the energetic analysis data are independent of any loudspeaker configuration. Applied to multichannel signals, the energetic sound field encoding can simultaneously provide lossy compression and means for automatic down- or upmixing to a chosen reproduction system. However, the analysis data often do not unambiguously determine the original signals. Depending both on the analyzed signals and on the loudspeaker configuration, there may be numerous possibilities to recreate the energetic sound field at the "sweet spot" or it may not be possible to achieve an exactly correct synthesis.

The work reported in [1, 2, 3, 4] relied on a psychoacoustically motivated formulation of the synthesis algorithm. This paper contributes to understanding of the underlying physical phenomena by studying how the signals of a stereo or multichannel loudspeaker system interact to create an energetic sound field, and what kinds of sound fields can be created by such systems. The paper derives fundamental analytical relationships between signal quantities and the energetic sound field quantities resulting from the reproduction of the corresponding signals. Furthermore, the shortcomings of the energetic encoding in representing the original signals are discussed.

The paper is organized as follows. Section 2 defines the scope of the analysis and the approximations involved therein, explains the notation used in the paper, and introduces the energetic analysis and the signal quantities of interest. Section 3 derives general relations between the signals of an arbitrary multichannel loudspeaker system and the reproduced energetic sound field. Section 4 elaborates on the illustrative case of a sound field produced by a pair of loudspeakers. Section 5 discusses the implications of the results on practical audio applications. Finally, conclusions are drawn in Section 6.

2. BACKGROUND AND DEFINITIONS

In order to simplify the sound field analysis, it is useful to make some assumptions about the reproduction of the loudspeaker signals. The assumptions are described below, followed by definition of the notation and the energetic and signal quantities of interest.

2.1. Limitations and approximations

All expressions in this paper are derived for analytical (complex) time-domain signals of arbitrary limited duration. The complex formulation enables applying the equations directly to individual transform indices (frequency bands) resulting from short-time Fourier analysis of the input signals. Moreover, the equations hold without modifications for real signals.

All loudspeakers are assumed to reproduce ideal plane waves in an anechoic environment with a sound pressure corresponding to the input signal. These assumptions were also implicitly used in the earlier application of the energetic sound field analysis to audio coding [4]. The plane wave approximation is valid in the far field of the loudspeakers. It considerably simplifies the derived equations and enables the analysis to be carried out independent of the absolute distances of the loudspeakers. All loudspeakers are further assumed to be equidistant from the analysis position and the common propagation delay is ignored in the equations. If desired, the transfer functions and propagation delays of individual loudspeakers could, however, be readily taken into account by applying them to each loudspeaker signal prior to performing the analysis.

The analysis will also be limited to discrete-time signals. This limitation is simply a matter of convenience; all of the expressions can also be formulated for continuous signals.

2.2. Vector notation

The following notational conventions are used throughout the paper. Time-domain signals are represented as column vectors and denoted with an arrow symbol ($\vec{\ }$). Spatial vectors in Cartesian coordinates are represented as row vectors and denoted with bold-face symbols. Furthermore, time-dependent spatial vectors are represented, combining both conventions from above, as matrices where each row corresponds to the spatial vector at one time instant. Alternatively, each column of these "spatial signal matrices" can be seen as a signal component.

Vector magnitudes are denoted with $\|\cdot\|$ and defined as follows:

signal vector $\vec{X} = [x_1 \ \ldots \ x_N]^T$:

$$\|\vec{X}\| = \left(\vec{X}^H \vec{X}\right)^{\frac{1}{2}} = \sqrt{\sum_{i=1}^{N} x_i^* x_i}\,, \tag{1}$$

spatial vector $\mathbf{X} = [x_1 \ \ldots \ x_M]$:

$$\|\mathbf{X}\| = \left(\mathbf{X}\mathbf{X}^H\right)^{\frac{1}{2}} = \sqrt{\sum_{i=1}^{M} x_i^* x_i}\,, \tag{2}$$

spatial signal matrix $\vec{\mathbf{X}} = [\vec{X}_1 \ \ldots \ \vec{X}_M]_{N\times M}$:

$$\|\vec{\mathbf{X}}\| = \sqrt{\sum_{i=1}^{M} \|\vec{X}_i\|^2} = \sqrt{\sum_{i=1}^{N} \|\mathbf{X}_i\|^2}\,, \tag{3}$$

where ^{T} denotes transposition, ^{H} denotes Hermitian transposition, and * denotes complex conjugation. In practice, $M = 2$ or $M = 3$ depending on whether the analysis is performed in two- or three-dimensional space.

2.3. Energetic quantities

The two key energetic quantities applied in [1, 2, 3, 4] were the active sound intensity and the energy density of the sound field. The active intensity vector describes the direction and amount of net transport of sound energy within an analysis time period[1] and can be computed as [6]

$$\mathbf{I}_a = \Re\left\{\vec{P}^H \vec{\mathbf{U}}\right\}, \tag{4}$$

where $\vec{P}$ is the sound pressure signal vector, $\vec{\mathbf{U}}$ is the particle velocity spatial signal vector, and $\Re\{\cdot\}$ denotes the real part of a complex number. Furthermore, the energy density of the sound field within the same analysis period is given by [6]

$$E = \frac{1}{2}\rho_0 \left[\frac{1}{Z_0^2}\|\vec{P}\|^2 + \|\vec{\mathbf{U}}\|^2\right], \tag{5}$$

where $Z_0 = \rho_0 c$ is the characteristic acoustic impedance of the air and ρ_0 is the mean density of the air and c is the speed of sound.

In the synthesis algorithm presented in [1, 2, 3, 4], the active intensity vector and the energy density were used to determine the direction of arrival of sound and the diffuseness of the sound field. The opposite of the direction of the active intensity vector was simply used as the direction of arrival of the net sound energy. Comparing the magnitude of the active intensity vector to the energy density, on the other hand, allowed estimating the diffuseness.[2] Specifically, Merimaa and Pulkki [1] defined the energetic diffuseness estimate as

$$\Psi = 1 - \frac{\|\mathbf{I}_a\|}{cE} = 1 - \frac{2Z_0\left\|\Re\left\{\vec{P}^H\vec{\mathbf{U}}\right\}\right\|}{\|\vec{P}\|^2 + Z_0^2\|\vec{\mathbf{U}}\|^2}. \tag{6}$$

This quantity describes the fraction of locally oscillating sound energy and can be shown to be bound between zero and one (for a related proof for the quantity $\|\mathbf{I}_a\|/E$ describing the velocity of the energy transport, see [8, 9]).

The sound pressure $\vec{P}$ and the particle velocity $\vec{\mathbf{U}}$ are related through the fluid momentum equation [6]. For a plane wave arriving from the direction of the unit vector $\mathbf{e}_u$ (propagating in the direction of $-\mathbf{e}_u$),

$$\vec{\mathbf{U}} = -\frac{1}{Z_0}\vec{P}\mathbf{e}_u, \tag{7}$$

where the result of the product $\vec{P}\mathbf{e}_u$ is a spatial signal matrix. Note that this relation does not hold for general sound fields but is valid under the adopted plane wave assumption.

2.4. Signal quantities

As will be shown in Section 3, the energetic sound field produced by an arbitrary configuration of equidistant loudspeakers can be expressed in terms of the directions of the loudspeakers and various fundamental signal quantities, which we define here for signals $\vec{X}_i$ and $\vec{X}_j$:

$$\text{correlation}: r_{ij} = \vec{X}_i^H\vec{X}_j = r_{ji}^*, \tag{8}$$

$$\text{correlation coefficient}: \phi_{ij} = \frac{r_{ij}}{\|\vec{X}_i\|\,\|\vec{X}_j\|} = \phi_{ji}^*, \tag{9}$$

$$\text{magnitude ratio}: m_{ij} = \frac{\|\vec{X}_i\|}{\|\vec{X}_j\|} = m_{ji}^{-1}. \tag{10}$$

In order to simplify the notation, we further define for N signals

$$\text{amplitude sum}: S_A = \sum_{i=1}^{N}\|\vec{X}_i\|, \tag{11}$$

$$\text{RMS sum}: S_{\text{RMS}} = \sqrt{\sum_{i=1}^{N}\|\vec{X}_i\|^2}. \tag{12}$$

Some of the quantities derived later can also be conveniently expressed in terms of what will be called

[1] Note that the usual definition of the active intensity is the *rate* of energy transport, i.e., the transport of energy per unit time [6] instead of within an arbitrary time window. However, the choice of a different temporal normalization in this paper simplifies the equations, and it is easy to show that it does not affect either the direction of the active intensity vector or the energetic diffuseness estimate.

[2] Strictly speaking, a typical definition of a diffuse sound field requires that all directions of propagation are equally probable and that the energy density is constant over the space of interest. A low active intensity compared to the energy density could be produced, for example, by the sound energy oscillating between two opposite directions regardless of whether sound propagation is occuring within other directions. It is thus a necessary but not a sufficient condition for a highly diffuse sound field. Nevertheless, the adopted measure has proven useful in audio applications. For some additional discussion, see [5, 7].

Gerzon vectors [10]. These vectors have been used in analysis of spatial sound reproduction techniques [11, 12] and have been applied to spatial audio coding by Goodwin and Jot [13, 14]. Consider N loudspeakers at the directions of the unit vectors $\mathbf{e}_i$ emitting signals $\vec{X}_i$, respectively. We define the Gerzon "velocity vector"[3] as

$$\mathbf{g}_V = \frac{1}{S_A}\sum_{i=1}^{N}\|\vec{X}_i\|\mathbf{e}_i \tag{13}$$

and the Gerzon "energy vector" as

$$\mathbf{g}_E = \frac{1}{S_{\mathrm{RMS}}^2}\sum_{i=1}^{N}\|\vec{X}_i\|^2\mathbf{e}_i\,. \tag{14}$$

Note that these Gerzon vectors generally point towards the direction of arrival of sound. The active intensity vector, on the other hand, points towards the direction of propagation, which results in negative signs in most of the relationships that are derived between these quantities in the following sections.

3. GENERAL RELATIONS

In this section, we derive general relationships between an arbitrary set of loudspeaker signals and the resulting sound field. As will be seen, simple relations can be found for the special cases of uncorrelated and fully positively correlated signals. For other correlation values, it is necessary to consider the interaction between each loudspeaker pair. A more detailed discussion is left to Section 4.

3.1. Active intensity

Consider again N loudspeakers at directions of the unit vectors $\mathbf{e}_i,\ i = 1\ldots N$ emitting signals $\vec{X}_i$, respectively. As described in Section 2.1, we define the sound pressure due to each loudspeaker at the analysis point as

$$\vec{P}_i = \vec{X}_i\,. \tag{15}$$

From Eqs. (4) and (7) the active intensity is now

$$\mathbf{I}_a = -\frac{1}{Z_0}\Re\left\{\sum_{i=1}^{N}\vec{X}_i^H\sum_{j=1}^{N}\vec{X}_j\mathbf{e}_j\right\}\,. \tag{16}$$

[3] The term velocity vector is somewhat misleading, since the nominator is proportional to the particle velocity of the sound field only if the loudspeaker signals are fully positively correlated.

By noting that $\vec{X}_i^H\vec{X}_j = r_{ij}$, the equation can be rearranged to

$$\begin{aligned}\mathbf{I}_a = -\frac{1}{Z_0}\Bigg(&\sum_{i=1}^{N}\|\vec{X}_i\|^2\mathbf{e}_i + \\ &+\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\Re\{r_{ij}\}(\mathbf{e}_i+\mathbf{e}_j)\Bigg)\,,\end{aligned} \tag{17}$$

where the relation $\Re\{r_{ij}\} = \Re\{r_{ji}\}$ has been used.

The second sum term in Eq. (17) shows that the closer two loudspeakers are to each other, the stronger the effect of the correlation of their signals is on the net flow of energy. Note also that it is the real part of the correlation that determines how it affects the energetic sound field. For the purposes of the energetic analysis, two signals with an imaginary correlation (two correlated signals with a 90° phase difference) thus appear equal to uncorrelated signals.

The first sum term in Eq. (17) can be identified as proportional to the Gerzon energy vector. For uncorrelated signals, the second sum term is zero and the active intensity can thus be written as

$$\mathbf{I}_{a\,\mathrm{uncorr}} = -\frac{1}{Z_0}S_{\mathrm{RMS}}^2\mathbf{g}_E\,. \tag{18}$$

That is, in this special case, the Gerzon energy vector fully indicates the directionality of the active intensity vector. An equally simple expression can also be found for the special case of fully positively correlated loudspeaker signals where $\vec{X}_i^H\vec{X}_j = \|\vec{X}_i\|\,\|\vec{X}_j\|$. From Eq. (16),

$$\mathbf{I}_{a\,\mathrm{corr}} = -\frac{1}{Z_0}S_A^2\mathbf{g}_V\,. \tag{19}$$

A simple general expression for the active intensity in terms of the previously defined Gerzon velocity and energy vectors does not seem to exist. However, the correlation terms can be manipulated to yield

$$\begin{aligned}\mathbf{I}_a = -\frac{1}{Z_0}\Bigg[&S_{\mathrm{RMS}}^2\mathbf{g}_E + \\ &+\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\Re\{\phi_{ij}\}\left(S_{Aij}^2\mathbf{g}_{Vij} - S_{\mathrm{RMS}ij}^2\mathbf{g}_{Eij}\right)\Bigg]\,,\end{aligned} \tag{20}$$

where $\mathbf{g}_{V\,ij}$ is the Gerzon velocity vector, $\mathbf{g}_{E\,ij}$ is the Gerzon energy vector, $S_{A\,ij}$ is the amplitude sum, and $S_{\mathrm{RMS}\,ij}$ is the RMS sum — all for the two loudspeaker signals $\vec{X}_i$ and $\vec{X}_j$.

3.2. Energy density and diffuseness

The energy density produced by the $\vec{X}_i$ loudspeaker signals is, from Eqs. (5) and (7),

$$E = \frac{\rho_0}{2Z_0^2}\left(\left\|\sum_{i=1}^{N}\vec{X}_i\right\|^2 + \left\|\sum_{i=1}^{N}\vec{X}_i\mathbf{e}_i\right\|^2\right) . \qquad (21)$$

This expression can also be simplified for the special cases of uncorrelated or fully positively correlated signals. For uncorrelated signals, $\left\|\sum_{i=1}^{N}\vec{X}_i\right\|^2 = \left\|\sum_{i=1}^{N}\vec{X}_i\mathbf{e}_i\right\|^2 = \sum_{i=1}^{N}\|\vec{X}_i\|^2$, so the energy density can be expressed as

$$E_{\mathrm{uncorr}} = \frac{\rho_0}{Z_0}S_{\mathrm{RMS}}^2 \, . \qquad (22)$$

For fully positively correlated signals, on the other hand, $\left\|\sum_{i=1}^{N}\vec{X}_i\right\|^2 = \left(\sum_{i=1}^{N}\|\vec{X}_i\|\right)^2$ and $\left\|\sum_{i=1}^{N}\vec{X}_i\mathbf{e}_i\right\|^2 = S_A^2\|\mathbf{g}_V\|^2$. Hence, the energy density can be expressed as

$$E_{\mathrm{corr}} = \frac{\rho_0}{2Z_0^2}S_A^2\left(1+\|\mathbf{g}_V\|^2\right) . \qquad (23)$$

Substituting Eqs. (18) and (22) into the definition of the diffuseness estimate Eq. (6), we now have

$$\Psi_{\mathrm{uncorr}} = 1 - \|\mathbf{g}_E\| \qquad (24)$$

and, correspondingly,

$$\Psi_{\mathrm{corr}} = \frac{\left(1-\|\mathbf{g}_V\|\right)^2}{1+\|\mathbf{g}_V\|^2} \, . \qquad (25)$$

4. STEREO REPRODUCTION

In this section, the general relationships derived above are applied to the special case of stereo loudspeaker reproduction, illustrating the effect of the correlation of the two loudspeaker signals on the energetic sound field.

4.1. Direction of the active intensity

Consider a pair of loudspeakers in the xy-plane at angles $\pm\theta_0$ relative to the x axis, emitting signals $\vec{X}_L$ and $\vec{X}_R$. The unit vectors pointing towards the loudspeakers are thus $\mathbf{e}_L = [\cos\theta_0 \;\; \sin\theta_0]$ and $\mathbf{e}_R = [\cos\theta_0 \;\; -\sin\theta_0]$. From Eq. (17), the x- and y-components of the active intensity vector are

$$I_{ax} = -\frac{\cos\theta_0}{Z_0}\left(\|\vec{X}_L\|^2 + \|\vec{X}_R\|^2 + 2\Re\{r_{LR}\}\right) , \qquad (26)$$

and

$$I_{ay} = -\frac{\sin\theta_0}{Z_0}\left(\|\vec{X}_L\|^2 - \|\vec{X}_R\|^2\right) , \qquad (27)$$

respectively.

The direction of arrival θ of the net sound energy (the opposite direction of the active intensity vector) can now be calculated from

$$\tan\theta = \frac{I_y}{I_x} = \frac{\|\vec{X}_L\|^2 - \|\vec{X}_R\|^2}{\|\vec{X}_L\|^2 + \|\vec{X}_R\|^2 + 2\Re\{r_{LR}\}}\tan\theta_0 \, . \qquad (28)$$

This equation does not depend on the absolute magnitudes of the input signals and can also be expressed in terms of the magnitude ratios m_{LR} and $m_{RL} = 1/m_{LR}$ of the signals as

$$\tan\theta = \frac{m_{LR} - m_{RL}}{m_{LR} + m_{RL} + 2\Re\{\phi_{LR}\}}\tan\theta_0 \, . \qquad (29)$$

The directions θ according to Eq. (29) are illustrated in Fig. 1. As can be expected, for an increasing magnitude ratio, the direction of arrival of the net sound energy asymptotically approaches that of the loudspeaker emitting the stronger signal. For equal magnitudes, on the other hand, the active intensity vector indicates sound energy arriving from halfway between the loudspeakers, apart from the case of fully negatively correlated signals ($\phi_{LR} = -1$) when the active intensity vector vanishes and the direction is not defined.

For small magnitude differences, the cross-correlation determines how much the weaker signal "pulls" the direction of arrival towards the corresponding loudspeaker. Furthermore, a negative correlation may even cause a "push" to beyond the directions spanned by the loudspeaker pair.[4] The reasons for this effect of the correlation

[4] The apparent direction of arrival of the sound energy from an angle beyond those spanned by the loudspeaker pair can be explained with circular flow of energy from one speaker to another. For further discussion, the interested reader is referred to [6].

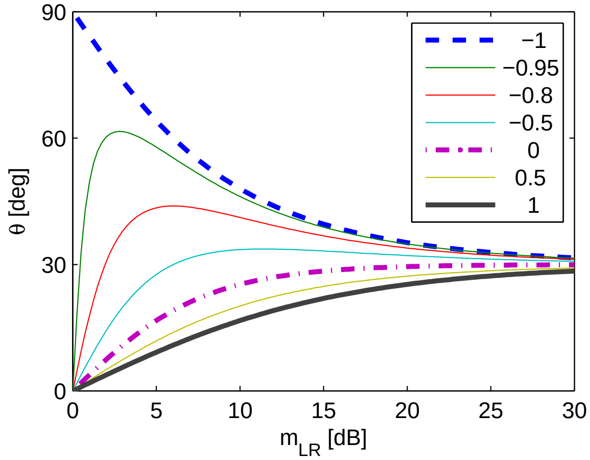

Fig. 1: Direction of arrival θ of the net sound energy as a function of the magnitude ratio m_{LR} for stereo loudspeakers at $\pm 30°$. The different curves correspond to different values of the cross-correlation coefficient ϕ_{LR}. Note that the intensity vector vanishes and the direction is not defined for $m_{LR} = 0$ dB and $\phi_{LR} = -1$.

can be easily seen from Eqs. (26) and (27). The y-component of the active intensity (parallel to the line connecting the loudspeakers) is independent of the correlation, whereas the x-component decreases as the correlation decreases, thus turning the direction of arrival of the sound energy away from the weaker loudspeaker.

From Eq. (28), the direction of arrival of the sound energy for fully positively correlated signals ($\Re\{r_{LR}\} = \|\vec{X}_L\| \, |\vec{X}_R\|$) is

$$\frac{\tan\theta_{\text{corr}}}{\tan\theta_0} = \frac{\|\vec{X}_L\| - \|\vec{X}_R\|}{\|\vec{X}_L\| + \|\vec{X}_R\|}. \tag{30}$$

This is the well-known tangent law of amplitude panning, which is widely used in audio technology to approximate the perceived direction of an auditory event [12, 15, 16]. For $\phi_{LR} = 0$, on the other hand, the direction of arrival of the sound energy corresponds to an "energy tangent law" (see [12])

$$\frac{\tan\theta_{\text{uncorr}}}{\tan\theta_0} = \frac{\|\vec{X}_L\|^2 - \|\vec{X}_R\|^2}{\|\vec{X}_L\|^2 + \|\vec{X}_R\|^2}. \tag{31}$$

From Eq. (20), the active intensity vector for reproduction with a pair of loudspeakers can also be written in terms of the Gerzon velocity and energy vectors as

$$\begin{aligned}\mathbf{I}_a = &-\frac{1}{Z_0}\Big[\Re\{\phi_{LR}\}\left(\|\vec{X}_L\| + \|\vec{X}_R\|\right)^2 \mathbf{g}_V + \\ &+ (1 - \Re\{\phi_{LR}\})\left(\|\vec{X}_L\|^2 + \|\vec{X}_R\|^2\right)\mathbf{g}_E\Big]\,.\end{aligned} \tag{32}$$

The real part of the cross-correlation coefficient can thus be interpreted as cross-fading between the Gerzon velocity and energy vectors (note the negative sign in the relationship between the directions of the Gerzon vectors and the active intensity vector). By considering Eq. (32) for fully positively correlated and uncorrelated signals, it is further obvious that the direction of the Gerzon velocity vector corresponds to the amplitude panning tangent law and the direction of the the Gerzon energy vector to the energy tangent law, as shown earlier in [12].

4.2. Diffuseness

From Eq. (21), the energy density for the stereo loudspeaker configuration can be shown to be

$$E = \frac{\rho_0}{Z_0^2}\left(\|\vec{X}_L\|^2 + \|\vec{X}_R\|^2 + 2\Re\{r_{LR}\}\cos^2\theta_0\right). \tag{33}$$

Furthermore, the magnitude of the active intensity vector is, from Eqs. (26) and (27),

$$\begin{aligned}\|\mathbf{I}_a\| = &\frac{1}{Z_0}\Big[\left(\|\vec{X}_L\|^2 - \|\vec{X}_R\|^2\right)^2 + \\ &+ 4\left(\|\vec{X}_L\|^2 + \Re\{r_{LR}\}\right)\times \\ &\times\left(\|\vec{X}_R\|^2 + \Re\{r_{LR}\}\right)\cos^2\theta_0\Big]^{\frac{1}{2}}\end{aligned} \tag{34}$$

Hence, the diffuseness estimate is as shown in Eq. (35).

The values of the diffuseness estimate as a function of the correlation coefficient and magnitude ratio of the input signals are illustrated in Fig. 2 for three different loudspeaker spacings: $\theta_0 = 30°$, $\theta_0 = 60°$, and $\theta_0 = 90°$. It can be seen that for a high magnitude difference, the diffuseness is always small, whereas for equal-magnitude and opposite-phase signals ($\phi_{LR} = -1,\ m_{LR} = 0$ dB), the sound field appears as fully diffuse regardless of the loudspeaker directions. For any intermediate values, the diffuseness estimate depends strongly on the angle between

$$\begin{aligned}\Psi &= 1 - \frac{\sqrt{\left(\|\vec{X}_L\|^2 - \|\vec{X}_R\|^2\right)^2 + 4\left(\|\vec{X}_L\|^2 + \Re\left\{r_{LR}\right\}\right)\left(\|\vec{X}_R\|^2 + \Re\left\{r_{LR}\right\}\right)\cos^2\theta_0}}{\|\vec{X}_L\|^2 + \|\vec{X}_R\|^2 + 2\Re\left\{r_{LR}\right\}\cos^2\theta_0} \\ &= 1 - \frac{\sqrt{\left(m_{LR} - m_{RL}\right)^2 + 4\left(m_{LR} + \Re\left\{\phi_{LR}\right\}\right)\left(m_{RL} + \Re\left\{\phi_{LR}\right\}\right)\cos^2\theta_0}}{m_{LR} + m_{RL} + 2\Re\left\{\phi_{LR}\right\}\cos^2\theta_0}\end{aligned} \tag{35}$$

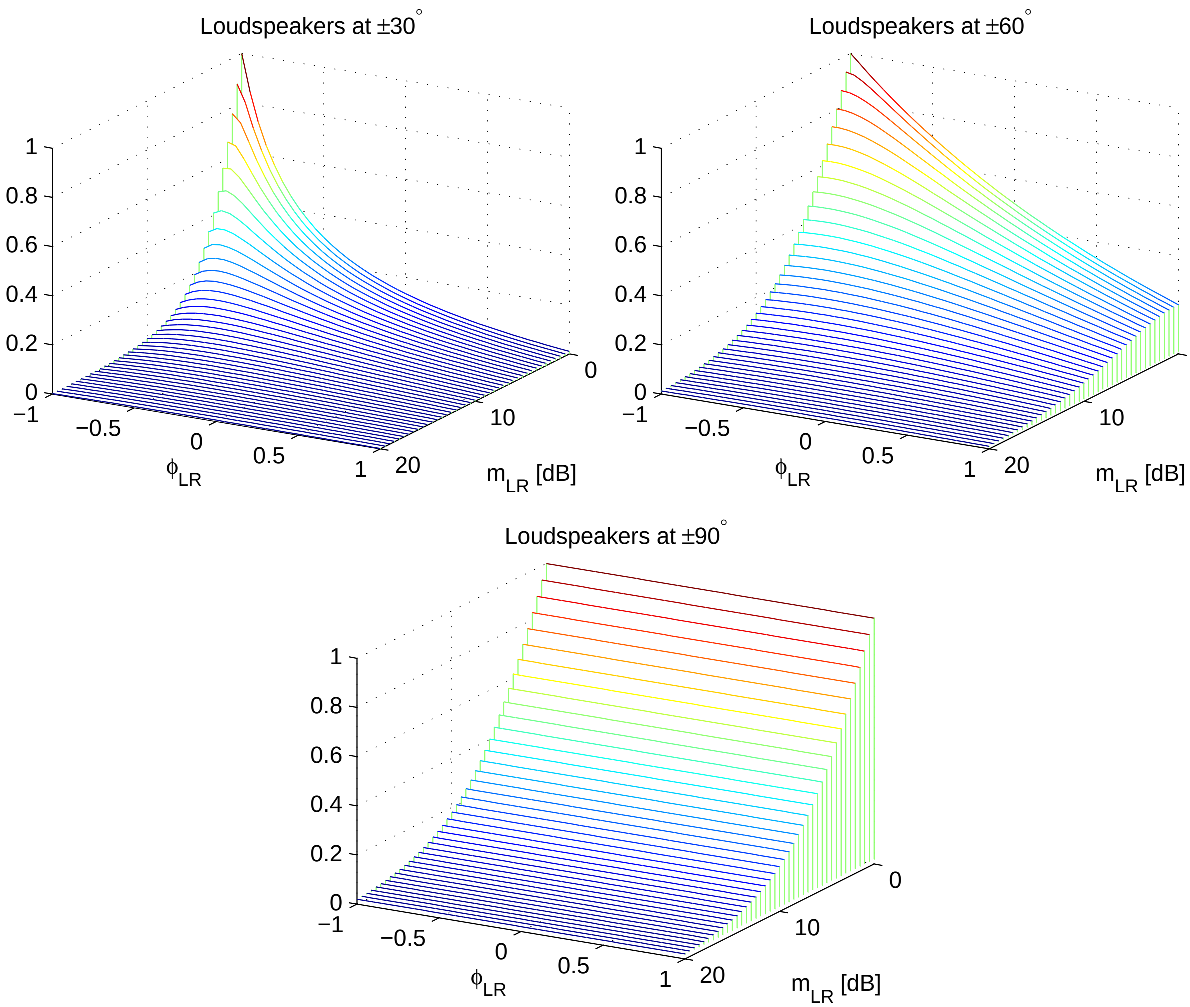

Fig. 2: Diffuseness Ψ as a function of the cross-correlation coefficient ϕ_{LR} and the magnitude ratio m_{LR} for stereo loudspeakers at $\pm 30°$, $\pm 60°$, and $\pm 90°$ azimuth.

the loudspeakers. For instance, the estimated diffuseness for uncorrelated equal-level signals reproduced with loudspeakers at $\pm 30°$ is smaller than that for fully positively correlated equal-level signals signals reproduced from $\pm 60°$. Furthermore, for loudspeakers at opposite directions, the estimated diffuseness is independent of the correlation and only depends on the magnitude ratio.

5. DISCUSSION

In prior methods where the energetic sound field quantities were decoded to synthesize loudspeaker signals [1, 2, 3, 4], the non-diffuse proportion of sound, as determined based on the diffuseness estimate, was reproduced as precisely as possible from the direction of arrival of the net sound energy. Based on the analysis in Sections 3 and 4, this approach may, however, yield unexpected results for signals with negative correlation: the analyzed direction of arrival may point to a direction outside the angle spanned by the loudspeakers originally reproducing the corresponding sound.[5] As shown in Section 4.2, negative correlation, on the other hand, tends to yield high estimated diffuseness. In the earlier energetic decoding, the diffuse part of sound was reproduced in a decorrelated form with all or most loudspeakers. Although not leading to a correct reproduction, such (partial) decorrelation might be a better compromise than a discrete rendering of the originally negatively correlated sound from a potentially abnormal direction.[6]

Potential problems also occur in encoding and decoding amplitude-panned sound. As shown in Secion 4.1, the estimated direction of arrival follows the tangent law and recreating it with amplitude panning is thus easy. However, the estimated diffuseness grows with an increasing distance between the loudspeakers used to reproduce the original signals. In decoding amplitude-panned signals to the same loudspeaker system as used originally, the resulting signals are obviously different from the originals if the diffuse part is reproduced as decorrelated. Both amplitude panning between widely spaced loudspeakers [18] and decorrelation [19] yield a perception of a broad auditory event. Hence, in upmixing to a loudspeaker system with speakers close to the analyzed direction, some of the broadness due to the original loudspeaker configuration is traded to broadness due to decorrelation. This might serve the purpose of the encoding and decoding. However, if the original intent was to create as narrow an auditory event as possible with the utilized loudspeaker system, the rendering based on the energetic quantities does not optimally "sharpen" the sound image.

So far, the discussion has concentrated on the limitations of the energetic encoding in capturing the spatial properties of the original signals. In addition, the analysis presented in this paper points out restrictions in the possibilities of rendering certain sound fields with a loudspeaker system. It is obviously impossible to render strictly non-diffuse sound from any direction between two loudspeakers (see Fig. 2). Furthermore, a highly diffuse sound field cannot be created with closely spaced loudspeakers without applying negative correlation.

It should be noted that high positive or negative correlations between several sound sources rarely appear in natural sound scenes. A discrete reflection from the acoustical environment may interact with the direct sound such that, depending on the delay of the reflection, the real part of the correlation fluctuates as a function of frequency. (The situation is identical to encoding and decoding delay-panned loudspeaker signals.) Nevertheless, as several reflections arrive to an analysis position, the true diffuseness of the sound field tends to increase in line with the energetic diffuseness estimate. The problems related to sources with high correlation thus mainly occur in encoding and decoding artificially created loudspeaker signals.

The shortcomings in the capability of the energetic analysis to capture the spatial properties of certain stereo or multichannel signals do not automatically mean that the analysis is not useful for the purposes of encoding multichannel signals. The derived relationships between the signal and the sound field quantities in stereo reproduction show that, given

[5] It is interesting to note the similarity to matrix decoding where an opposite phase between the signals (which translates to negative correlation) signifies sound to be rendered behind the listener.

[6] Whether or not analyzed directions of arrival outside those subtended by a loudspeaker pair are abnormal, is also debatable. In fact, opposite phase is applied in virtual loudspeaker technology to specifically extend the sound image to directions not subtended by a pair of loudspeakers [17]. In the future, it would be interesting to study the active intensity resulting from such sound reproduction.

the directions of the loudspeakers used in the encoding, it would be possible to revise the synthesis such that the magnitude ratio and the real part of the cross-correlation coefficient of the synthesized signals are identical to those of the original signals. In multichannel encoding, information is lost in the conversion to sound field quantities, but this is the case in any form of lossy audio coding. More work will be needed to assess the merits and drawbacks of the energetic approach compared to signal-based approaches in encoding of typical multichannel content.

6. CONCLUSIONS

It was shown that the active intensity, energy density, and estimated diffuseness resulting from loudspeaker reproduction of stereo and multichannel signals can be expressed in terms of the signal magnitudes, cross-correlations, and loudspeaker directions. The analysis revealed that for fully positively correlated and uncorrelated signals, the energetic quantities are proportional to the Gerzon velocity and energy vectors, respectively. Furthermore, the direction of arrival of the net sound energy resulting from amplitude panning between two loudspeakers corresponds to the tangent law of amplitude panning.

A number of potential problems related to applying energetic analysis as a way to encode stereo or multichannel sound were also identified. In the investigation of stereo reproduction, it was shown that negative correlation between two loudspeaker signals may lead to a net flow of energy that appears to be emanating from an angle not subtended by the loudspeaker pair. Although the estimated diffuseness of the sound field increases with a decreasing correlation between two loudspeaker signals, the estimate is also highly sensitive to the angular separation between the utilized loudspeakers. Specifically, amplitude panning between widely spaced speakers may yield a sound field with a high estimated diffuseness. Further work will be needed to assess the severeness of these problems in practical applications.

7. ACKNOWLEDGEMENTS

The author would like to thank Mike Goodwin and Jean-Marc Jot for initiating this study and for subsequent discussions and comments.

8. REFERENCES

[1] J. Merimaa and V. Pulkki, "Spatial Impulse Response Rendering I: Analysis and synthesis," *J. Audio Eng. Soc.*, vol. 53, no. 12, pp. 1115–1127, 2005.

[2] V. Pulkki and J. Merimaa, "Spatial Impulse Response Rendering II: Reproduction of diffuse sound and listening tests," *J. Audio Eng. Soc.*, vol. 54, no. 1/2, pp. 3–20, 2006.

[3] V. Pulkki and J. Merimaa, "Spatial Impulse Response Rendering: Listening tests and applications to continuous sound," in *AES 118th Convention*, 2005. Preprint 6371.

[4] V. Pulkki, "Spatial sound reproduction with Directional Audio Coding," *J. Audio Eng. Soc.*, vol. 55, no. 6, pp. 503–516, 2007.

[5] A. Hurtado-Huyssen and J.-D. Polack, "Acoustic intensity in multichannel rendering systems," in *AES 119th Convention*, 2005. Preprint 6548.

[6] F. J. Fahy, *Sound Intensity.* Essex, England: Elsevier Science Publishers Ltd., 1989.

[7] J. Merimaa, *Analysis, Synthesis, and Perception of Spatial Sound — Binaural Localization Modeling and Multichannel Loudspeaker Reproduction.* PhD thesis, Helsinki University of Technology, 2006. Available at http://lib.tkk.fi/Diss/2006/isbn9512282917/.

[8] G. Schiffrer and D. Stanzial, "Energetic properties of acoustic fields," *J. Acoust. Soc. Am.*, vol. 96, no. 4, pp. 3645–3653, 1994.

[9] D. Stanzial, N. Prodi, and G. Schiffrer, "Reactive acoustic intensity for general fields and energy polarization," *J. Acoust. Soc. Am.*, vol. 99, no. 4, pp. 1868–1876, 1996.

[10] M. A. Gerzon, "General metatheory of auditory localisation," in *AES 92nd Convention*, 1992. Preprint 3306.

[11] M. A. Gerzon and G. J. Barton, "Ambisonic decoders for HDTV," in *AES 92nd Convention*, 1992. Preprint 3345.

[12] J.-M. Jot, V. Larcher, and J.-M. Pernaux, "A comparative study of 3-D audio encoding and rendering techniques," in *AES 16th International Conference*, pp. 281–300, 1999.

[13] M. M. Goodwin and J.-M. Jot, "A frequency-domain framework for spatial audio coding based on universal spatial cues," in *AES 120th Convention*, 2006. Preprint 6751.

[14] M. M. Goodwin and J.-M. Jot, "Analysis and synthesis for universal spatial audio coding," in *AES 121st Convention*, 2006. Preprint 6874.

[15] J. C. Bennett, K. Barker, and F. O. Edeko, "A new approach to the assessment of stereophonic sound system performance," *J. Audio Eng. Soc.*, vol. 33, no. 5, pp. 314–321, 1985.

[16] V. Pulkki, "Virtual sound source positioning using vector base amplitude panning," *J. Audio Eng. Soc.*, vol. 45, no. 6, pp. 456–466, 1997.

[17] D. H. Cooper and J. L. Bauck, "Prospects for transaural recording," *J. Audio Eng. Soc.*, vol. 37, no. 1/2, pp. 3–19, 1989.

[18] V. Pulkki and M. Karjalainen, "Localization of amplitude-panned virtual sources I: Stereophonic panning," *J. Audio Eng. Soc.*, vol. 49, no. 9, pp. 739–752, 2001.

[19] K. Kurozumi and K. Ohgushi, "The relationship between the cross-correlation coefficient for two-channel acoustic signals and sound image quality," *J. Acoust. Soc. Am.*, vol. 74, no. 6, pp. 1726–1733, 1983.