GEOMETRIC SIGNAL DECOMPOSITIONS FOR SPATIAL AUDIO ENHANCEMENT

Michael M. Goodwin

Creative Advanced Technology Center
Scotts Valley, CA
mgoodwin@atc.creative.com

ABSTRACT

Decomposition of audio signals into primary and ambient components is useful for realizing spatial enhancements such as upmix and stereo widening. In this paper, we present several methods for primary-ambient decomposition of two-channel audio signals based on signal-space geometry. We discuss the performance of the various methods with respect to target orthogonality conditions on the estimated primary and ambient components, which cannot all be satisfied due to the need to constrain the model components to the signal subspace in light of limitations on implementation complexity.

***Index Terms*—** spatial audio, primary-ambient decomposition, signal analysis, signal representations

1. INTRODUCTION

In a variety of spatial audio analysis-synthesis applications, it is useful or even necessary to separate the input audio signal into primary and ambient components for processing and/or rendering; these applications include upmix [1, 2, 3], multichannel format conversion, stereo widening, headphone reproduction [4], and spatial audio coding [5]. In processing primary components, which correspond to discrete sources, the spatial properties should be preserved or stabilized. For example, in a stereo reproduction, the spatial position of phantom sources collapses to the nearest loudspeaker; in 2-to-5 upmix, *i.e.* expanding a stereo signal for multichannel reproduction, the frontal imaging can be maintained for a wider listening area by populating the center channel with appropriate primary content extracted from the stereo input [2]. For ambient components such as reverberation, applause, or rain, the goal in spatial enhancement algorithms is typically to achieve a perceptual impression of envelopment, so such components should be rendered with an appropriate spatial diffuseness [6].

This paper is organized as follows. Section 2 further discusses the motivation for primary-ambient decomposition, describes prior methods based on scalar time-frequency masks, and proposes a vector decomposition model and orthogonality constraints for the model. Several vector decomposition methods for primary-ambient separation are considered in Section 3; these are based generally on orthogonal projections, principal components analysis, and signal-dependent bases. Concluding remarks are given in Section 4.

2. PRIMARY-AMBIENT SIGNAL DECOMPOSITION

In its simplest form, a primary-ambient decomposition of a stereo signal can be expressed as

$$\vec{x}_L = \vec{p}_L + \vec{a}_L \quad (1)$$
$$\vec{x}_R = \vec{p}_R + \vec{a}_R \quad (2)$$

where $\vec{x}_L$ and $\vec{x}_R$ are the left and right channels of the stereo signal, $\vec{p}_L$ and $\vec{p}_R$ are the respective primary components, and $\vec{a}_L$ and $\vec{a}_R$ are the corresponding ambient components. The vectors $\vec{x}_L$ and $\vec{x}_R$ here could either be the original time-domain audio signals or subband signals in a time-frequency representation, where the latter case is typically preferable in that it provides some separation of the signal components. Given the primary-ambient signal model of (1), then, the task is to estimate the primary and ambient components for each channel signal. The general idea in the model estimation is that primary components in the two channels should be highly correlated (except for the case where a primary source is hard-panned, *i.e.* present in only one of the channels) and that the ambient components in the two channels should be uncorrelated; furthermore, the primary and ambient components within a single channel should be uncorrelated as well. These assumptions about the correlation properties stem from concepts in psychoacoustics (in that perception of diffuseness is related to interaural signal decorrelation), room acoustics (in that late reverberation at different points in a room tends to be uncorrelated), and in studio recording practices (wherein uncorrelated stereo reverb is often added in the production process) [6, 7].

2.1. Scalar mask methods

In several methods for ambience extraction described in the literature, the ambience components are estimated by applying a scalar mask to the corresponding channel signal:

$$\vec{a}_L = A_L \vec{x}_L \quad (3)$$
$$\vec{a}_R = A_R \vec{x}_R \quad (4)$$

where the masks are based on the channel signal auto-correlations and/or cross-correlation [2, 8, 9]. The primary components are then given simply by

$$\vec{p}_L = (1 - A_L)\vec{x}_L \quad (5)$$
$$\vec{p}_R = (1 - A_R)\vec{x}_R. \quad (6)$$

In such decompositions, it is clear that the correlation coefficient between the estimated components (either primary or ambient) is the same as that between the original channel signals. For the case where $\vec{x}_L$ and $\vec{x}_R$ are the original time-domain signals, such a scalar mask approach clearly undermines the target inter-component correlation conditions described earlier. Where $\vec{x}_L$ and $\vec{x}_R$ constitute subband signals in a time-frequency representation of the input, the correlation conditions of course do not hold on a per-subband basis; however, a trend toward meeting the correlation relationships can be observed for the time-domain signals generated from the estimated subband primary and ambient components, especially if the primary and ambient components are well resolved in the time-frequency representation [9].

2.2. Vector decompositions

In order to improve the performance of primary-ambient decompositions for spatial audio applications, we consider in this paper various estimation approaches which, unlike scalar mask methods, satisfy at least some of the target correlation conditions directly in the decomposition. The basic idea is to derive primary and ambient unit vectors for each channel such that the model in (1) can be further specified as:

$$\vec{x}_L = \rho_L \vec{v}_L + \alpha_L \vec{e}_L \quad (7)$$

$$\vec{x}_R = \rho_R \vec{v}_R + \alpha_R \vec{e}_R. \quad (8)$$

where $\vec{v}_L$ and $\vec{v}_R$ are the primary unit vectors, $\vec{e}_L$ and $\vec{e}_R$ are the ambience unit vectors, and where the expansion coefficients ρ_L, ρ_R, α_L, and α_R describe the level and balance of the components.

Ideally, according to the assumptions discussed earlier, the unit vectors should satisfy the constraints:

$$\vec{v}_L = \vec{v}_R \quad (9)$$

$$\vec{v}_L^H \vec{e}_L = 0 \quad (10)$$

$$\vec{v}_R^H \vec{e}_R = 0 \quad (11)$$

$$\vec{e}_L^H \vec{e}_R = 0 \quad (12)$$

such that the primary components constitute a common fully correlated source and the various inter-component orthogonality conditions are satisfied. In the first condition, we are essentially assuming that only a single primary source is active in the two-channel signal; in this light, carrying out such decompositions on the subband signals in a time-frequency representation (such as the short-time Fourier transform) is advantageous in that this source assumption is more likely to be valid on a per-subband basis than for the original time-domain signals.

Given that the signals $\vec{x}_L$ and $\vec{x}_R$ define a two-dimensional signal space, it is necessary to consider directions outside of the signal subspace if the three orthogonality conditions (10)-(12) are to be met. This excursion is problematic both in that the decomposition problem is then under-specified and in that the complexity is prohibitive for practical applications in consumer audio devices. For the purposes of this paper, then, we restrict the considerations to unit component vectors in the signal subspace, *i.e.* we use decomposition vectors which can be derived as a linear combination of the original signal vectors. As will be discussed in Section 3, some of the constraints must be relaxed given this restriction.

3. GEOMETRIC DECOMPOSITIONS

Signal-space geometry provides a useful visualization of signal decompositions in that the correlation relationships between the various components are immediately evident. In this section, we discuss several decompositions based on signal-space geometry, focusing on which of the constraints in (9)-(12) are satisfied by the respective approaches. As will become clear, the various approaches are fundamentally defined by how the unit vectors in the primary-ambient signal model are determined.

3.1. Cross-channel projection

The *cross-channel projection* (CCP) approach is based initially on the premise that the ambience component in a given channel should be uncorrelated with the full signal in the other channel. If the orthogonality conditions in (10)-(12) are indeed met, we have, *e.g.*

$$\vec{e}_L^H \vec{x}_R = \vec{e}_L^H (\vec{v}_R + \vec{e}_R) = \vec{e}_L^H \vec{v}_R + \vec{e}_L^H \vec{e}_R = 0. \quad (13)$$

While the conditions

$$\vec{e}_L^H \vec{x}_R = 0 \qquad \vec{e}_R^H \vec{x}_L = 0 \quad (14)$$

of course do not imply that the component orthogonality conditions in (10)-(12) are necessarily satisfied, they do provide a useful basis for establishing the decomposition vectors. Initially, suppose we select

$$\vec{v}_L = \frac{\vec{x}_R}{\|\vec{x}_R\|} \qquad \vec{v}_R = \frac{\vec{x}_L}{\|\vec{x}_L\|}. \quad (15)$$

Then, the conditions in (14) are met if ρ_L and ρ_R are determined by orthogonal projection:

$$\rho_L = \vec{v}_L^H \vec{x}_L \qquad \rho_R = \vec{v}_R^H \vec{x}_R. \quad (16)$$

The ambience components are then given by the projection residuals:

$$\alpha_L \vec{e}_L = \vec{x}_L - (\vec{v}_L^H \vec{x}_L)\vec{v}_L \quad (17)$$

$$\alpha_R \vec{e}_R = \vec{x}_R - (\vec{v}_R^H \vec{x}_R)\vec{v}_R \quad (18)$$

which yields the decomposition shown in Figure 1(a), where the dashed lines indicate the ambience components.

The CCP representation is useful for some stereo enhancement effects in that it characterizes the inter-channel signal differences in the directions $\vec{e}_L$ and $\vec{e}_R$, but it has a clear shortcoming in the primary component estimation in that $\vec{v}_L$ and $\vec{v}_R$ are no more correlated than the original signals. To address this, consider modifying the decomposition in Figure 1(a) by reallocating some of the estimated ambience component to the primary component for each of the channels:

$$\vec{x}_L = (\rho_L \vec{v}_L + \beta_L \vec{e}_L) + (\alpha_L - \beta_L)\vec{e}_L \quad (19)$$

$$\vec{x}_R = (\rho_R \vec{v}_R + \beta_R \vec{e}_R) + (\alpha_R - \beta_R)\vec{e}_R. \quad (20)$$

The effect of this reallocation is that the ambience unit vectors are preserved, but the primary unit vectors are modified so as to "focus" the primary components in the decomposition. Of course, there are infinitely many solutions for the adjustment gains $\{\beta_L, \beta_R\}$ such that the modified primary components are fully correlated (colinear). Two such solutions are depicted in Figure 1(b) and Figure 1(c), based respectively on the assumptions that the two channels have equal ambient-to-signal energy ratios and that the ambience components in the two channels have equal energy [9]. The latter assumption is typically favorable in practice, as the ambience is generally balanced between the two channels in stereo recordings; the former assumption tends to break down in the presence of a primary source that is dominant in one of the channels.

It should be clear from the illustrations in Figure 1 that the modified CCP decomposition satisfies the primary component constraint in (9), but does not actually meet any of the orthogonality constraints in (10)-(12). On the other hand, it should also be clear that the method provides a reasonable relaxation of the orthogonality constraints in that none of the constraints are radically violated. Indeed, the adjustment gains $\{\beta_L, \beta_R\}$ could be determined so as to balance the component correlations subject to some optimization criterion, but such an extension is beyond the scope of this paper.

3.2. Principal components analysis

In [10], a primary-ambient decomposition method based on principal components analysis (PCA) is presented. In the algorithm, the principal component vector is found as

$$\vec{d} = (\vec{x}_L^H \vec{x}_R)\vec{x}_L + (\lambda_0 - \vec{x}_L^H \vec{x}_L)\vec{x}_R \quad (21)$$

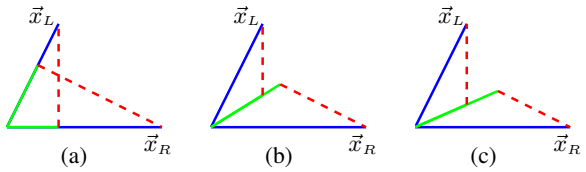

Fig. 1. Signal decomposition using cross-channel projection and modification: (a) orthogonal projection; (b) with equal ambience ratios; (c) with equal ambience energies. The ambience components are shown as dashed lines.

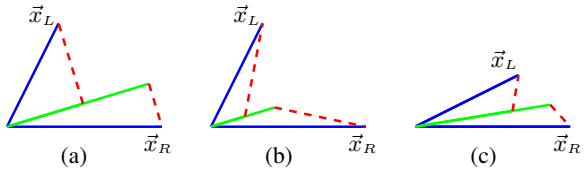

Fig. 2. Primary-ambient decomposition using principal components analysis: (a) PCA; (b) the PCA decomposition in (a) with an *ad hoc* modification to improve the decomposition of uncorrelated inputs; (c) an example of the modified PCA decomposition for a more strongly correlated signal.

where λ_0 is the dominant eigenvalue of the input covariance matrix. Based on an assumption that the primary component constitutes the majority of the energy in the audio signal, the primary unit vector is selected as

$$\vec{v} = \frac{\vec{d}}{\|\vec{d}\|} \tag{22}$$

and the ambience components are estimated as the residuals after orthogonal projection of the channel signals onto $\vec{v}$:

$$\alpha_L \vec{e}_L = \vec{x}_L - (\vec{v}^H \vec{x}_L)\vec{v} \tag{23}$$
$$\alpha_R \vec{e}_R = \vec{x}_R - (\vec{v}^H \vec{x}_R)\vec{v}. \tag{24}$$

This approach finds the unit vector $\vec{v}$ that best describes the input signals: it maximizes the sum of the projection energies $|\vec{v}^H \vec{x}_L|^2 + |\vec{v}^H \vec{x}_R|^2$ and likewise minimizes the residual energy $|\alpha_L|^2 + |\alpha_R|^2$. As such, if the primary component is dominant in the input signal, the principal PCA component provides a robust estimate of this primary component.

An example of a PCA-based decomposition is shown in Figure 2(a). The PCA decomposition satisfies the primary commonality constraint (9) and the primary-ambient orthogonality conditions (10)-(11) by construction. However, the constraint (12) is violated in that the estimated ambience components are actually colinear (with a negative correlation). Furthermore, when the input signals are not highly correlated (and the primary dominance assumption does not hold), the PCA approach overestimates the primary component in the decomposition [9]. While the PCA method provides a perceptually compelling primary component for many natural audio signals, it is necessary to address these shortcomings in a general algorithm. In the following two sections, corrective methods which leverage the PCA primary component estimation but improve the decomposition for weakly correlated signals are described.

3.3. Modified PCA

The PCA-based primary-ambient decomposition relies on the assumption that the primary component is dominant. When this is the case, as in many audio recordings, the primary component extraction is perceptually compelling. However, as shown in [9], the PCA underestimates the amount of ambience energy, most markedly when the two channels are uncorrelated (and there is no true primary component); instead of identifying both channels as ambient, it selects the higher-energy channel as the principal component (which corresponds to the primary unit vector in the decomposition) and the lower-energy channel as the secondary component (which corresponds to the ambience unit vector). The PCA is thus clearly valid only when the dominance assumption holds, *i.e.* when $|\phi_{LR}|$ is close to one. As $|\phi_{LR}|$ approaches zero, the primary-ambient decomposition would indeed be better estimated by considering the signal to be entirely ambient. This observation suggests an *ad hoc* modification of the PCA decomposition:

$$\vec{x}_L = |\phi_{LR}|\left(\rho_L \vec{v}_L + \alpha_L \vec{e}_L\right) + \left(1 - |\phi_{LR}|\right)\vec{x}_L \tag{25}$$
$$= |\phi_{LR}|\rho_L \vec{v}_L + |\phi_{LR}|\alpha_L \vec{e}_L + \left(1 - |\phi_{LR}|\right)\vec{x}_L \tag{26}$$
$$\vec{x}_R = \underbrace{|\phi_{LR}|\rho_R \vec{v}_R}_{} + \underbrace{|\phi_{LR}|\alpha_R \vec{e}_R + \left(1 - |\phi_{LR}|\right)\vec{x}_R}_{} \tag{27}$$

where the underbraces indicate the modified primary and ambient components. An example of this modified PCA decomposition is depicted in Figure 2(b), where it should be clear that the estimated ambience components are significantly less correlated that in the PCA decomposition of Figure 2(a). Informal listening tests indicate that this approach provides a substantial improvement over PCA for synthetic test signals and a slight improvement for typical music audio.

3.4. Orthogonal ambience basis expansion

Of the methods described in the previous sections, none provide a decomposition that explicitly satisfies the inter-channel ambience orthogonality condition in (12). In this section, we develop a method that ensures the ambience components are always orthogonal by directly constructing the ambience unit vectors to be orthogonal, *i.e.* to constitute an orthonormal basis for the signal subspace. The basis is derived such that

$$\frac{\vec{e}_L^H \vec{x}_L}{\|\vec{x}_L\|} = \frac{\vec{e}_R^H \vec{x}_R}{\|\vec{x}_R\|} \tag{28}$$

which ensures that the ambience basis functions are not biased with respect to either of the input signals. Furthermore, if the input signals are fully uncorrelated, the ambience unit vectors will be found as normalized versions of the signals themselves.

The ambience basis derivation consists of two steps: first, an orthogonal basis for the signal subspace is constructed using a Gram-Schmidt process:

$$\vec{g}_L = \frac{\vec{x}_L}{\|\vec{x}_L\|} \tag{29}$$
$$\vec{g}_R = \vec{x}_R - (\vec{g}_L^H \vec{x}_R)\vec{g}_L \tag{30}$$

where $\vec{g}_R$ is subsequently normalized. Then, the ambience unit vectors are determined by rotating the Gram-Schmidt basis:

$$\begin{bmatrix} \vec{e}_L & \vec{e}_R \end{bmatrix} = \frac{1}{(1 + |\gamma|^2)^{\frac{1}{2}}} \begin{bmatrix} \vec{g}_L & \vec{g}_R \end{bmatrix} \begin{bmatrix} 1 & -\gamma^* \\ \gamma & 1 \end{bmatrix} \tag{31}$$

where

$$\gamma = \frac{1}{\phi_{LR}} \left[-1 + \left(1 - |\phi_{LR}|^2\right)^{\frac{1}{2}} \right] \tag{32}$$

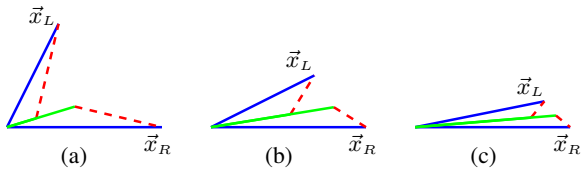

Fig. 3. Primary-ambient decomposition using a signal-adaptive orthogonal ambience basis and a primary unit vector derived by PCA.

is used; this choice of γ rotates the Gram-Schmidt basis such that the resulting ambience unit vectors $\vec{e}_L$ and $\vec{e}_R$ satisfy the condition in (28).

After the ambience basis is derived, each channel is decomposed using the corresponding ambience unit vector and a primary unit vector derived via PCA; we retain the PCA unit vector in this algorithm due to its robust performance for correlated (*i.e.* mostly primary) input signals. The expansion coefficients are given by

$$\begin{bmatrix} \rho_L \\ \alpha_L \end{bmatrix} = \left(\begin{bmatrix} \vec{v} & \vec{e}_L \end{bmatrix}^H \begin{bmatrix} \vec{v} & \vec{e}_L \end{bmatrix} \right)^{-1} \begin{bmatrix} \vec{v} & \vec{e}_L \end{bmatrix}^H \vec{x}_L \quad (33)$$

$$\begin{bmatrix} \rho_R \\ \alpha_R \end{bmatrix} = \left(\begin{bmatrix} \vec{v} & \vec{e}_R \end{bmatrix}^H \begin{bmatrix} \vec{v} & \vec{e}_R \end{bmatrix} \right)^{-1} \begin{bmatrix} \vec{v} & \vec{e}_R \end{bmatrix}^H \vec{x}_R \quad (34)$$

which can be readily simplified as

$$\rho_L = \frac{\vec{v}^H \vec{x}_L - (\vec{v}^H \vec{e}_L)(\vec{e}_L^H \vec{x}_L)}{1 - |\vec{v}^H \vec{e}_L|^2} \quad (35)$$

$$\alpha_L = \frac{\vec{e}_L^H \vec{x}_L - (\vec{e}_L^H \vec{v})(\vec{v}^H \vec{x}_L)}{1 - |\vec{v}^H \vec{e}_L|^2} \quad (36)$$

and similarly for ρ_R and α_R. If the input signals are not correlated, the ambience basis expansion coefficients α_L and α_R will be dominant, whereas if the input signals are highly correlated, the primary coefficients will be dominant. We can thus view this as a formalization of the modification in Section 3.3, with the distinction that the ambience component orthogonality is always ensured here. Several examples of signal decomposition using this orthogonal ambience basis approach are illustrated in Figure 3; note that the ambience components are orthogonal in all cases. This ambience orthogonality leads to an improved subjective quality for the decomposition with respect to the other methods described in this paper, at the cost of the additional computation needed to derive the ambience basis.

4. CONCLUSIONS AND FUTURE WORK

In this paper we have presented several vector-space methods for decomposing a two-channel audio signal into primary and ambient components. The methods were discussed with respect to desired orthogonality conditions for the model components, and graphical depictions of the decompositions were used to illustrate the performance with respect to these conditions. Future work includes extending these methods to the multichannel case, wherein a robust primary-ambient decomposition of an arbitrary number of input channels is needed.

5. REFERENCES

[1] R. Irwan and R. M. Aarts, "Two-to-five channel sound processing," *J. Audio Eng. Soc.*, vol. 50, no. 11, pp. 914–926, Nov. 2002.

[2] C. Avendano and J.-M. Jot, "A frequency-domain approach to multichannel upmix," *J. Audio Eng. Soc.*, vol. 52, no. 7/8, pp. 740–749, July/Aug. 2004.

[3] C. Faller, "Multiple-loudspeaker playback of stereo signals," *J. Audio Eng. Soc.*, vol. 54, no. 11, pp. 1051–1064, Nov. 2006.

[4] M. Goodwin and J.-M. Jot, "Binaural 3-D audio rendering based on spatial audio scene coding," *123rd Audio Eng. Soc. Conv.*, Oct. 2007, Preprint 7277.

[5] M. Goodwin and J.-M. Jot, "A frequency-domain framework for spatial audio coding based on universal spatial cues," *120th Audio Eng. Soc. Conv.*, May 2006, Preprint 6751.

[6] G. Kendall, "The decorrelation of audio signals and its impact on spatial imagery," *Computer Music Journal*, vol. 19, no. 4, pp. 71–87, Winter 1995.

[7] J. A. Moorer, "About this reverberation business," *Computer Music Journal*, vol. 3, no. 2, pp. 13–28, June 1979.

[8] C. Avendano and J.-M. Jot, "Ambience extraction and synthesis from stereo signals for multi-channel audio up-mix," *IEEE-ICASSP*, vol. 2, pp. 1957–1960, May 2002.

[9] J. Merimaa, M. Goodwin, and J.-M. Jot, "Correlation-based ambience extraction from stereo recordings," *123rd Audio Eng. Soc. Conv.*, Oct. 2007, Preprint 7282.

[10] M. Goodwin and J.-M. Jot, "Primary-ambient decomposition and vector-based localization for spatial audio coding and analysis-synthesis," *IEEE-ICASSP*, vol. 1, pp. 9–12, April 2007.