Paper published in IEEE/ACM Trans. Audio, Speech, Lang. Processing, vol. 22. no. 2, pp. 505-517, Feb. 2014

http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6698287

Linear Estimation Based Primary-Ambient Extraction for Stereo Audio Signals

Jianjun He, *Student Member*, *IEEE*, Ee-Leng Tan, and Woon-Seng Gan, *Senior Member*, *IEEE*

Abstract

Audio signals for moving pictures and video games are often linear combinations of primary and ambient components. In spatial audio analysis-synthesis, these mixed signals are usually decomposed into primary and ambient components to facilitate flexible spatial rendering and enhancement. Existing approaches such as principal component analysis (PCA) and least squares (LS) are widely used to perform this decomposition from stereo signals. However, the performance of these approaches in primary-ambient extraction (PAE) has not been well studied and no comparative analysis among the existing approaches has been carried out so far. In this paper, we generalize the existing approaches into a linear estimation framework. Under this framework, we propose a series of performance measures to identify the components that contribute to the extraction error. Based on the generalized linear estimation framework and our proposed performance measures, a comparative study and experimental testing of the linear estimation based PAE approaches including existing PCA, LS, and three proposed variant LS approaches are presented.

Index Terms

Primary-ambient extraction (PAE), spatial audio, linear estimation, principal component analysis (PCA), least squares (LS), performance measure

This work is supported by the Singapore Ministry of Education Academic Research Fund Tier-2, under research grant MOE2010-T2-2-040. Digital Object Identifier 10.1109/TASLP.2013.2297015. Jianjun He, Ee-Leng Tan and Woon-Seng Gan are with Digital Signal Processing Lab, School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, 639798(Email: jhe007@e.ntu.edu.sg, etanel@ntu.edu.sg, ewsgan@ntu.edu.sg).

Paper published in IEEE/ACM Trans. Audio, Speech, Lang. Processing, vol. 22. no. 2, pp. 505-517, Feb. 2014

http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6698287

Linear Estimation Based Primary-Ambient Extraction for Stereo Audio Signals

I. INTRODUCTION

With the increasing prevalence of 3D video technology, consumers are demanding a more immersive listening experience to better match the 3D visual effects. This results in a growing need for a better spatial audio reproduction. In moving pictures and video games, audio signals generally consist of point-like sound sources and environmental sound, which shall be referred to as primary and ambient components, respectively [1], [2]. To achieve accurate rendering of spatial audio, different processing schemes should be applied to the primary and ambient components of the audio signals [2], [3]. However, the primary and ambient components in moving pictures and video games are mixed in stereo and multichannel signals [4], which necessitates primary-ambient extraction (PAE). Fig. 1 shows the extraction of the primary and ambient components from a stereo signal using PAE. Recent years have seen applications of PAE in spatial audio processing [3], [5]-[9], spatial audio coding [8], [10], [11], audio up-mixing [1], [9], [12], [13], and immersive 3D sound systems [14]-[16].

In addition to binaural cue coding (BCC) [17] and MPEG surround [18], there are two emerging frameworks in spatial audio coding, namely, spatial audio scene coding (SASC) [8], [11] and directional audio coding (DirAC) [10]. Both SASC and DirAC aim to reproduce spatial sound using any sound system configurations, though DirAC is mainly targeted for acoustic signals. One essential stage of these methods is to separate the input audio signal into primary (or non-diffuse) and ambient (or diffuse) sound. In SASC, the localization analysis using the Gerzon localization vector [19] is carried out separately for the decomposed primary and ambient components and the spatial cues are applied in the final synthesis. In

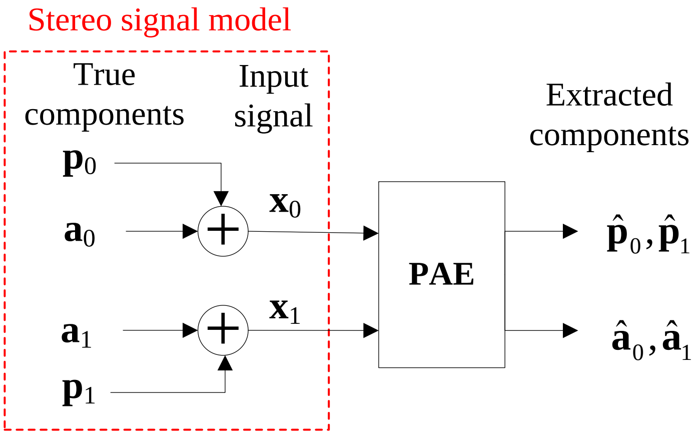


Fig. 1. Extraction of the primary and ambient components using PAE, where $\mathbf{x}_0, \mathbf{x}_1$ are the input stereo signals; $\mathbf{p}_0, \mathbf{p}_1$ and $\mathbf{a}_0, \mathbf{a}_1$ are the true primary and ambient components; $\hat{\mathbf{p}}_0, \hat{\mathbf{p}}_1$ and $\hat{\mathbf{a}}_0, \hat{\mathbf{a}}_1$ are the extracted primary and ambient components.

DirAC, the primary sound is reproduced using vector base amplitude panning (VBAP) [20], while the ambient sound is usually decorrelated to create a better surround sound effect.

Various up-mixing techniques based on PAE have been discussed in [1], [9], [13]. The PAE based up-mixing approaches are particularly suitable for the immersive 3D (i3D) sound system proposed by Gan *et al.* [14], [15]. The i3D sound system comprises a unique combination of conventional and parametric loudspeakers, which aims to accurately reproduce the primary and ambient components of the spatial sound so as to deliver an immersive and realistic soundscape for gaming and home entertainment applications [16]. This sound system exploits the high directivity of the parametric loudspeaker to render sharp images of the primary components, as well as the conventional loudspeaker to reproduce the spaciousness of the ambient components.

PAE can also be applied to other research problems in audio signal processing. Similar to PAE, blind source separation (BSS) deals with the decomposition of significantly different components from the mixed signals. The key difference between BSS and PAE is the characteristics of the separated components: BSS separates the physical source components [21], while PAE extracts the components based on their inter-channel relationship, which is related to the perceptual spatial features. One of the most important features of PAE is the extraction of accurate spatial cues from the input signal, which is in line with the objectives in sound localization problems. By considering the primary and ambient components as the direct and reverberant sound, respectively, PAE can be applied to extract the reverberant sound [1], [9], [12], [13], [22]. PAE can also be applied to noise reduction if the extraction of the primary component from a noise-like ambient component is considered [23].

To date, many approaches have been proposed for PAE from stereo signals, and PAE has been extended to deal with multichannel signals [2], [24]. For these approaches dealing with stereo signals, the audio signal is generally modeled as directional sound sources mixed with diffuse ambient sound. In such a stereo signal model, the key difference between the primary and ambient components is discriminated by their correlations between the two channels, i.e., the primary and ambient components are considered to be correlated and uncorrelated, respectively [2]. In [1], a time-frequency mask is used to extract the ambient component from a stereo signal assuming that the ambient component is of equal level in the two channels or equal ratio to primary component in the two channels. In [25], Faller introduced a least squares (LS) approach to estimate the primary and ambient components for surround sound up-mixing. Principal component analysis (PCA) remains one of the most widely studied approaches applied in PAE [2], [16], [26]-[33]. Considering the independence between the primary and ambient components, two orthogonal bases can be obtained using the Karhunen-Loève transform from the signal space of the stereo signal [34]. Based on the assumption that the primary component is relatively stronger than the ambient component, the projected signal on the basis vector with larger variance is assumed to be the primary component, and the projected signal on the other basis vector is assumed as the ambient component. Experimental studies in [27] show that PCA and LS based PAE produce superior extraction results than the time-frequency masks, especially in primary extraction. Even though PCA and LS based PAE are popular approaches, the relationship and differences between PCA and LS based PAE as well as the performance of PCA and LS in PAE still remain unclear. Other techniques applied in PAE include non-negative matrix factorization [35] and independent component analysis [21], [36].

In this paper, we focus on PAE approaches based on linear estimation, which assume that the primary and ambient components are linearly mixed in the stereo signal model [2]. Based on the linear estimation, PCA and LS are designed to minimize the correlation between the primary and ambient components and the extraction error, respectively. Our analysis reveals that the extraction error consists of three error components, namely, distortion, interference, and leakage. Distortion relates to the amount of amplitude scaling of the extracted primary (or ambient) component as compared to the true primary (or ambient) component. Interference measures the amount of uncorrelated primary (or ambient) component that is extracted from the stereo signal. Leakage measures the amount of undesired ambient (or primary) components in the extracted primary (or ambient) component. The characteristics of these three error components indicate that the leakage and distortion are perceptually more influential than interference in most of the applications. Taking this into consideration, different solutions for PAE can be obtained by minimizing these components. By minimizing the leakage and distortion, two variant LS approaches, namely, minimum leakage LS (MLLS) and minimum distortion LS (MDLS) are proposed in this paper, respectively. This derivation is followed by a comparative study on the performance of these PAE approaches. Based on our observations of this comparison, another approach referred to as the adjustable LS (ALS) is proposed, which offers adjustable error performance between the distortion and extraction error. Four major contributions of this paper are summarized as follows:

1) Proposed a linear estimation framework for PAE;

2) Suggested two groups of measures to give a more complete performance evaluation of the PAE approaches;
3) Proposed three PAE approaches, namely, MLLS, MDLS, and ALS, and conducted a comprehensive evaluation and comparison of these approaches together with the existing PAE approaches;
4) Provided practical guidelines in selecting the proper PAE approaches in spatial audio applications.

The rest of this paper is organized as follows. In Section II, we review the stereo signal model, and the key assumptions of this signal model. Subsequently, the linear estimation framework of PAE and two groups of performance measures are presented in Section III. Section IV discusses several approaches applied in PAE. Section V presents our discussion on the simulation results, which leads to our recommendations in applying the PAE approaches in different applications. Section VI concludes this work.

II. STEREO SIGNAL MODEL

Sound scenes in moving pictures and video games usually comprise several point-like sound sources (or primary component) and the environmental ambient sound (or ambient component) [4]. PAE aims to separate the primary component from the ambient component based on their perceptual spatial features. The perceptual spatial features can be characterized by the inter-channel relationships, including inter-channel time difference (ICTD), inter-channel level difference (ICLD), and inter-channel cross-correlation coefficient (ICC) [17]. Since the number of primary sources is usually unknown and might be varying, a common practice in spatial audio processing is to convert the signals into time-frequency domain using short-time Fourier transform (STFT) [1], [2], [8], [10], [25], [27], [37] or subband via filter banks like hybrid quadrature mirror filter banks [18]. For each frequency band or subband, it is generally assumed that the primary component of the input signal is composed of only one dominant source [1], [2], [25], [27].

Paper published in IEEE/ACM Trans. Audio, Speech, Lang. Processing, vol. 22. no. 2, pp. 505-517, Feb. 2014

http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6698287

Denoting the *m*th subband of input stereo signals at time index *l* as $\mathbf{x}_0[m,l]=[x_0(0),\ldots,x_0(N-1)]^T$, and $\mathbf{x}_1[m,l]=[x_1(0),\ldots,x_1(N-1)]^T$, where *N* is the length of one frame. PAE is carried out in each subband of each frame independently, and the extracted primary and ambient components are combined via inverse STFT or synthesis filter banks. The stereo signal model is expressed as:

$$\begin{aligned}\mathbf{x}_0[m,l]&=\mathbf{p}_0[m,l]+\mathbf{a}_0[m,l],\\ \mathbf{x}_1[m,l]&=\mathbf{p}_1[m,l]+\mathbf{a}_1[m,l],\end{aligned} \tag{1}$$

where $\mathbf{p}_0,\mathbf{p}_1$ and $\mathbf{a}_0,\mathbf{a}_1$ are the primary and ambient components in the two channels of the stereo signal, respectively. Since the subbands of the input signal are exclusively used in the analysis of PAE approaches, the indices $[m,l]$ are omitted for brevity.

The stereo signal model also assumes the primary and ambient components in the two channels to be correlated and uncorrelated, respectively. The correlation coefficient between the two channels of the signal $\mathbf{x}_i$ and $\mathbf{x}_j$ is defined as $\phi_{ij}(\tau)=r_{ij}(\tau)\big/\sqrt{r_{ii}(0)r_{jj}(0)}$, where $r_{ij}(\tau)$ is the correlation between $\mathbf{x}_i$ and $\mathbf{x}_j$ at lag τ. Two signals are considered correlated when $\max_{\tau}|\phi_{ij}(\tau)|=1;$ uncorrelated when $\max_{\tau}|\phi_{ij}(\tau)|=0;$ and partially correlated when $0<\max_{\tau}|\phi_{ij}(\tau)|<1.$

Correlated primary component in the stereo signal can be described by one of the following conditions [38]: i) amplitude panned, i.e., $\mathbf{p}_1=k\mathbf{p}_0,$ where *k* is referred to as the primary panning factor (PPF); ii) time shifted, i.e., $p_1(n)=p_0(n+\tau),$ where $p_1(n)$ is the *n*th sample of $\mathbf{p}_1$ and τ is the ICTD; and iii) amplitude panned and time shifted, i.e., $p_1(n)=kp_0(n+\tau).$ In this signal model, we only consider the primary component to be amplitude panned by PPF *k* [2], [25], [27]. This amplitude panned primary component is commonly found in stereo recordings using pan pot stereo and coincident techniques as well as sound mixes using panning [4]. For an ambient component that consists of environmental sound, it is usually considered to be uncorrelated with the primary component [22], [39], [40]. The ambient component in the two channels is also assumed to be uncorrelated and relatively balanced in terms of power, considering the diffuseness of ambient component. To quantify the power difference between the primary and ambient components, we introduce the primary power ratio (PPR) γ, which is defined as the ratio of total primary power to total signal power in two channels:

$$\gamma = \left(P_{\mathbf{p}_0} + P_{\mathbf{p}_1}\right) \Big/ \left(P_{\mathbf{x}_0} + P_{\mathbf{x}_1}\right), \tag{2}$$

where $P_{(.)}$ denotes the mean square power of the signal in the subscript. From (2), it is clear that γ ranges from zero to one. Summarizing the assumptions for the stereo signal model, we have

$$\mathbf{p}_1 = k\mathbf{p}_0,\ \mathbf{a}_0 \perp \mathbf{a}_1,\ \mathbf{p}_i \perp \mathbf{a}_j, \forall i, j \in \{0,1\}, \tag{3}$$

$$P_{\mathbf{p}_1} = k^2 P_{\mathbf{p}_0},\ \ P_{\mathbf{a}_1} = P_{\mathbf{a}_0}, \tag{4}$$

where $\perp$ represents that two signals are uncorrelated.

Given any stereo input signal that fulfills the above conditions, the relationships between the auto- and cross-correlations at zero-lag and the power of these components can be expressed as

$$r_{00} = {\mathbf{x}_0}^H \mathbf{x}_0 = NP_{\mathbf{x}_0} = N\left(P_{\mathbf{p}_0} + P_{\mathbf{a}_0}\right), \tag{5}$$

$$r_{11} = {\mathbf{x}_1}^H \mathbf{x}_1 = NP_{\mathbf{x}_1} = N\left(k^2 P_{\mathbf{p}_0} + P_{\mathbf{a}_0}\right), \tag{6}$$

$$r_{01} = {\mathbf{x}_0}^H \mathbf{x}_1 = {\mathbf{p}_0}^H \mathbf{p}_1 = NkP_{\mathbf{p}_0}, \tag{7}$$

where H is the Hermitian transpose operator. From (5)-(7), the PPF and PPR of the stereo signal are

$$k = \frac{r_{11} - r_{00}}{2r_{01}} + \sqrt{\left(\frac{r_{11} - r_{00}}{2r_{01}}\right)^2 + 1}, \tag{8}$$

$$\gamma = \frac{2r_{01} + \left(r_{11} - r_{00}\right)k}{\left(r_{11} + r_{00}\right)k}. \tag{9}$$

The primary component is panned to channel 1 for $k > 1$ and to channel 0 for $k < 1$. In spatial audio, the PPF is considered as the square root of ICLD. Only the primary or ambient component is found in the stereo signal for $\gamma = 1$ or $\gamma = 0$, respectively. In other words, the primary component becomes more prominent as γ increases. In the following sections, we shall see that PPF and PPR are useful parameters for the extraction of the primary and ambient components, as well as to evaluate the performance of the PAE approaches.

III. LINEAR ESTIMATION FRAMEWORK AND PERFORMANCE MEASURES

In this paper, we examine the blind extraction of primary and ambient components from a stereo input signal. Inspired by the mixing signal model given in (1), we address the PAE problem based on a linear estimation framework, where the primary and ambient components are estimated as weighted sums of the stereo signals in two channels. Thus, the extracted primary and ambient components are expressed as

$$\begin{bmatrix} \hat{\mathbf{p}}_0^T \\ \hat{\mathbf{p}}_1^T \\ \hat{\mathbf{a}}_0^T \\ \hat{\mathbf{a}}_1^T \end{bmatrix} = \begin{bmatrix} w_{\mathrm{P}0,0} & w_{\mathrm{P}0,1} \\ w_{\mathrm{P}1,0} & w_{\mathrm{P}1,1} \\ w_{\mathrm{A}0,0} & w_{\mathrm{A}0,1} \\ w_{\mathrm{A}1,0} & w_{\mathrm{A}1,1} \end{bmatrix} \begin{bmatrix} \mathbf{x}_0^T \\ \mathbf{x}_1^T \end{bmatrix} = \mathbf{W} \begin{bmatrix} \mathbf{x}_0^T \\ \mathbf{x}_1^T \end{bmatrix}, \tag{10}$$

where $\hat{\mathbf{p}}_0, \hat{\mathbf{p}}_1$ and $\hat{\mathbf{a}}_0, \hat{\mathbf{a}}_1$ are the extracted primary and ambient components in the two channels, respectively; T is the transpose operator; and $w_{(.)}$ is the estimated weight of the extracted component, where the first subscript "P" or "A" denotes the primary or ambient component, respectively, the second subscript denotes the channel of the extracted component, and the third subscript denotes the channel of the input signal. Using this formulation, the PAE problem is simplified to the estimation of weighting matrix $\mathbf{W}$.

Based on the weighting matrix $\mathbf{W}$, we shall introduce two groups of measures to evaluate the objective performance of the linear estimation based PAE approaches. The first group measures the extraction accuracy of the primary and ambient components, whereas the second group examines the accuracy of the localization cues for the primary component and diffuseness for the ambient component.

A. Group 1: Measures for Extraction Accuracy

In [27], the extraction accuracy of PAE approaches is evaluated by the similarity measures based on the cross-correlation coefficient between the extracted and true components. While these measures quantify the overall performance of the PAE approaches, these measures are unable to provide in-depth insights on possible causes for the performance degradation. In this subsection, we shall analyze the components that form the extraction error of the PAE approaches, and propose four performance measures to quantify the extraction error. A similar decomposition on the error components with corresponding measures can be found in [41]. In the following, we discuss the error measures for the primary component first and then for the ambient component.

Considering the error between the extracted primary component $\hat{\mathbf{p}}_0$ and its true component $\mathbf{p}_0$, we have

$$\boldsymbol{\varepsilon}_{\mathrm{P}} = \hat{\mathbf{p}}_0 - \mathbf{p}_0. \tag{11}$$

Based on (11), we compute the error-to-signal ratio (ESR) for the primary component, which is defined as the ratio of the power of the extraction error to the power of the true primary component:

$$\mathrm{ESR}_{\mathrm{P}} = P_{\boldsymbol{\varepsilon}_{\mathrm{P}}} \big/ P_{\mathbf{p}_0}. \tag{12}$$

Note that the ESR is equivalent to the normalized mean square error (NMSE).

Based on (10), $\hat{\mathbf{p}}_0$ can be expressed as

$$\hat{\mathbf{p}}_0 = w_{\mathrm{P}0,0}\mathbf{x}_0 + w_{\mathrm{P}0,1}\mathbf{x}_1. \tag{13}$$

Based on the assumptions stated in (3) and substituting (1) into (13), we have

$$\begin{aligned}\hat{\mathbf{p}}_0 &= \left(w_{\mathrm{P}0,0}\mathbf{p}_0 + w_{\mathrm{P}0,1}\mathbf{p}_1\right) + \left(w_{\mathrm{P}0,0}\mathbf{a}_0 + w_{\mathrm{P}0,1}\mathbf{a}_1\right)\\ &= w_{\mathrm{P}0}\mathbf{p}_0 + \left(w_{\mathrm{P}0,0}\mathbf{a}_0 + w_{\mathrm{P}0,1}\mathbf{a}_1\right)\\ &= \mathbf{p}_0 + \left(w_{\mathrm{P}0} - 1\right)\mathbf{p}_0 + \left(w_{\mathrm{P}0,0}\mathbf{a}_0 + w_{\mathrm{P}0,1}\mathbf{a}_1\right),\end{aligned} \tag{14}$$

where $w_{\mathrm{P}0} = w_{\mathrm{P}0,0} + kw_{\mathrm{P}0,1}$ is the weight of $\mathbf{p}_0$ in the extracted component $\hat{\mathbf{p}}_0$. Substituting (14) into (11), the extraction error becomes

$$\boldsymbol{\varepsilon}_{\mathrm{P}} = \left(w_{\mathrm{P}0} - 1\right)\mathbf{p}_0 + \left(w_{\mathrm{P}0,\,0}\mathbf{a}_0 + w_{\mathrm{P}0,\,1}\mathbf{a}_1\right) = Dist_{\mathrm{P}} + Leak_{\mathrm{P}}, \tag{15}$$

where $Dist_{\mathrm{P}} = \left(w_{\mathrm{P}0} - 1\right)\mathbf{p}_0$ and $Leak_{\mathrm{P}} = w_{\mathrm{P}0,0}\mathbf{a}_0 + w_{\mathrm{P}0,1}\mathbf{a}_1$ are the distortion and leakage in the extraction error, respectively. The distortion comes from the extraction weight $w_{\mathrm{P}0}$, which fluctuates from frame to frame, causing variations in sound timbre or level. We consider the primary component to be completely extracted and hence distortionless when $w_{\mathrm{P}0} = 1$. On the other hand, the leakage of the extracted primary component $Leak_{\mathrm{P}}$ originates from the true ambient components $\mathbf{a}_0$ and $\mathbf{a}_1$ of the stereo signal. We consider the ratios of the distortion and leakage power to the power of true primary component, as the distortion-to-signal ratio (DSR) [23] and the leakage-to-signal ratio (LSR), respectively:

$$\begin{aligned}\mathrm{DSR}_{\mathrm{P}} &= P_{Dist_{\mathrm{P}}} \big/ P_{\mathbf{p}_0},\\ \mathrm{LSR}_{\mathrm{P}} &= P_{Leak_{\mathrm{P}}} \big/ P_{\mathbf{p}_0}.\end{aligned} \tag{16}$$

Similar performance measures are also obtained to quantify the ambient extraction error. Based on (10), the extraction error of the ambient component is rewritten as

$$\begin{aligned}\boldsymbol{\varepsilon}_{\mathrm{A}} &= \hat{\mathbf{a}}_0 - \mathbf{a}_0\\ &= \left(w_{\mathrm{A}0,0} - 1\right)\mathbf{a}_0 + w_{\mathrm{A}0,1}\mathbf{a}_1 + \left(w_{\mathrm{A}0,0}\mathbf{p}_0 + w_{\mathrm{A}0,1}\mathbf{p}_1\right)\\ &= Dist_{\mathrm{A}} + Intf_{\mathrm{A}} + Leak_{\mathrm{A}},\end{aligned} \tag{17}$$

where the three components in $\boldsymbol{\varepsilon}_{\mathrm{A}}$: $Dist_{\mathrm{A}} = \left(w_{\mathrm{A}0,0} - 1\right)\mathbf{a}_0$, $Intf_{\mathrm{A}} = w_{\mathrm{A}0,1}\mathbf{a}_1$, and $Leak_{\mathrm{A}} = w_{\mathrm{A}0,0}\mathbf{p}_0 + w_{\mathrm{A}0,1}\mathbf{p}_1$ are the distortion, interference, and leakage, respectively. Similar to primary extraction, the distortion comes from the extraction weight $w_{\mathrm{A}0,0}$, and the ambient component is considered to be distortionless when $w_{\mathrm{A}0,0} = 1$. Interference $Intf_{\mathrm{A}}$ is produced by the uncorrelated ambient component in the counterpart channel $\mathbf{a}_1$, whereas the leakage of the extracted ambient component $Leak_{\mathrm{A}}$ originates from true primary components $\mathbf{p}_0$ and $\mathbf{p}_1$. The extraction error of the ambient component and its three error components are quantified by the ratios of their power to the power of true ambient component, as ESR, DSR, interference-to-signal ratio (ISR), and LSR, which are given as

$$\begin{aligned}
\mathrm{ESR}_{\mathrm{A}} &= P_{\boldsymbol{\varepsilon}_{\mathrm{A}}} \big/ P_{\mathbf{a}_0}, \\
\mathrm{DSR}_{\mathrm{A}} &= P_{Dist_{\mathrm{A}}} \big/ P_{\mathbf{a}_0}, \\
\mathrm{ISR}_{\mathrm{A}} &= P_{Intf_{\mathrm{A}}} \big/ P_{\mathbf{a}_0}, \\
\mathrm{LSR}_{\mathrm{A}} &= P_{Leak_{\mathrm{A}}} \big/ P_{\mathbf{a}_0}.
\end{aligned} \tag{18}$$

Comparing the measures of extraction error for the primary and ambient components, we find that no interference is found in the extracted primary component due to the unity correlation of the primary component. For both the primary and ambient components, ESR quantifies the overall error of the extracted component, and DSR, ISR, LSR provide detailed information on the extraction performance. In particular, LSR corresponds to the perceptual difference between the primary and ambient components. Both the interference and distortion in the extracted primary (or ambient) component come from the differences in this primary (or ambient) component between the two channels, hence they often exhibit some perceptual similarity with the true primary (or ambient) component. However, leakage solely comes from the ambient (or primary) component. Consequently, leakage is much more noticeable and undesirable than interference and distortion. On this note, we consider LSR to be the most important measure among DSR, ISR, and LSR for many applications. Nevertheless, more emphasis should be placed on DSR when sound timbre or level is of high importance.

B. Group 2: Measures for Spatial Accuracy

In the second group of measures, we consider the spatial accuracy of the extracted primary component based on three widely used spatial cues, namely, ICC, ICTD, and ICLD. These cues are used to evaluate the sound localization accuracy of the extracted primary component [5], [38]. There have been many studies to estimate ICTD after the coincidence model proposed by Jeffress (see [42]-[43] and references therein). Based on the Jeffress model [42], the ICC at different time lags is calculated and the lag number corresponds to the maximum ICC is the estimated ICTD. ICLD is obtained by taking the ratio of the power between the signals in two channels.

As the ambient component is assumed to be uncorrelated and balanced in the two channels, ICC and ICLD are selected as the measures to determine the diffuseness of the extracted ambient component [44]. A better extraction of the ambient component is obtained when the ICC and ICLD of the extracted ambient component are closer to zero and one, respectively.

IV. PRIMARY-AMBIENT EXTRACTION BASED ON LINEAR ESTIMATION

Following the discussions in Section III, we shall derive the solutions for PAE approaches using linear estimation. These solutions are obtained by optimizing the weights in **W** for different criteria in PAE, including the minimization of the correlation between primary and ambient components, and the minimization of different error components. In this section, an analytic study and comparison of five linear estimation based PAE approaches including three proposed approaches will be presented.

A. Primary-Ambient Extraction Using Principal Component Analysis

Principal component analysis is a widely used method in multivariate analysis [34]. The central idea of PCA is to linearly transform its input sequence into orthogonal principal components with descending variances. PCA was first introduced to solve the PAE problem in [28], and a closed-form solution of PCA based PAE for stereo signals can be obtained by eigenvalue decomposition of the input covariance matrix [2].

In general, the primary component is assumed to possess more power than the ambient component, i.e., $\gamma > 0.5$. Hence, it is a common practice to relate the larger eigenvalue to the primary component and the smaller eigenvalue to the ambient component. First, we find the larger eigenvalue and its corresponding primary basis vector [2], [27] with

$$\lambda_{\mathrm{P}} = 0.5\left[r_{00} + r_{11} + \sqrt{\left(r_{00} - r_{11}\right)^2 + 4r_{01}{}^2}\right], \tag{19}$$

$$\mathbf{u}_{\mathrm{P}} = r_{01}\mathbf{x}_0 + \left(\lambda_{\mathrm{P}} - r_{00}\right)\mathbf{x}_1. \tag{20}$$

Next, we compute the extracted primary components as

$$\begin{aligned}\hat{\mathbf{p}}_{\mathrm{PCA},0} &= \frac{\mathbf{u}_{\mathrm{P}}{}^H\mathbf{x}_0}{\mathbf{u}_{\mathrm{P}}{}^H\mathbf{u}_{\mathrm{P}}}\mathbf{u}_{\mathrm{P}}, \\ \hat{\mathbf{p}}_{\mathrm{PCA},1} &= \frac{\mathbf{u}_{\mathrm{P}}{}^H\mathbf{x}_1}{\mathbf{u}_{\mathrm{P}}{}^H\mathbf{u}_{\mathrm{P}}}\mathbf{u}_{\mathrm{P}}.\end{aligned} \tag{21}$$

Using (5)-(9), the expressions for the extracted primary components using PCA are simplified to (detailed derivation can be found in the appendix)

$$\begin{aligned}\hat{\mathbf{p}}_{\mathrm{PCA},0} &= \frac{1}{1+k^2}\left(\mathbf{x}_0 + k\mathbf{x}_1\right), \\ \hat{\mathbf{p}}_{\mathrm{PCA},1} &= \frac{k}{1+k^2}\left(\mathbf{x}_0 + k\mathbf{x}_1\right) = k\hat{\mathbf{p}}_{\mathrm{PCA},0}.\end{aligned} \tag{22}$$

Similarly, the extracted ambient components are obtained as

$$\begin{aligned}\hat{\mathbf{a}}_{\text{PCA},0} &= \frac{k}{1+k^2}\left(k\mathbf{x}_0 - \mathbf{x}_1\right),\\ \hat{\mathbf{a}}_{\text{PCA},1} &= -\frac{1}{1+k^2}\left(k\mathbf{x}_0 - \mathbf{x}_1\right) = -\frac{1}{k}\hat{\mathbf{a}}_{\text{PCA},0}.\end{aligned} \tag{23}$$

From (22)-(23), we observe that the weights for the extracted primary and ambient components are solely dependent on *k*. Between the two channels, the primary components are amplitude panned by a factor of *k*, whereas the ambient components are negatively correlated and panned to the opposite direction of the primary components, as indicated by the scaling factor $-1/k$. Clearly, the assumption of the uncorrelated ambient components in the stereo signal model does not hold considering the ambient components extracted using PCA. This drawback is inevitable in PCA since the ambient components in two channels are obtained from the same basis vector. As the primary and ambient components are derived from different basis vectors, the assumption that the primary components are uncorrelated with the ambient components is well satisfied in PCA.

By substituting the true primary and ambient components into (22) and (23), we have

$$\begin{aligned}\hat{\mathbf{p}}_{\text{PCA},0} &= \mathbf{p}_0 + \frac{1}{1+k^2}\left(\mathbf{a}_0 + k\mathbf{a}_1\right),\\ \hat{\mathbf{p}}_{\text{PCA},1} &= \mathbf{p}_1 + \frac{k}{1+k^2}\left(\mathbf{a}_0 + k\mathbf{a}_1\right),\end{aligned} \tag{24}$$

$$\begin{aligned}\hat{\mathbf{a}}_{\text{PCA},0} &= \frac{k^2}{1+k^2}\mathbf{a}_0 - \frac{k}{1+k^2}\mathbf{a}_1,\\ \hat{\mathbf{a}}_{\text{PCA},1} &= \frac{1}{1+k^2}\mathbf{a}_1 - \frac{k}{1+k^2}\mathbf{a}_0.\end{aligned} \tag{25}$$

Since there is no primary component in (25), (25) or (23) which comes from the basis vector with the smaller eigenvalue cannot be related with the extraction of the primary components. That is to say, the basis vector with larger eigenvalue always corresponds to the primary component regardless of the value of the primary power ratio γ. This observation reveals that the assumption $\gamma > 0.5$ in PCA is redundant.

However, if this assumption is not satisfied in the stereo input signal, the extraction error of the extracted primary component becomes higher, as inferred from (24).

B. Primary-Ambient Extraction Using Least Squares

Least squares estimation is frequently used to approximate solutions for over-determined systems. According to the stereo signal model, Faller introduced LS to extract the primary and ambient components by minimizing the MSE of the extracted components [25]. Considering the extraction of the primary component, the extraction error expressed in (15) can then be rewritten as

$$\boldsymbol{\varepsilon}_{\mathrm{P}} = \hat{\mathbf{p}}_0 - \mathbf{p}_0 = \left(w_{\mathrm{P}0,0} + k w_{\mathrm{P}0,1} - 1\right)\mathbf{p}_0 + w_{\mathrm{P}0,0}\mathbf{a}_0 + w_{\mathrm{P}0,1}\mathbf{a}_1, \tag{26}$$

and the MSE is $J = E\left[\boldsymbol{\varepsilon}_{\mathrm{P}}{}^{H}\boldsymbol{\varepsilon}_{\mathrm{P}}\right]$. By substituting the assumptions and relationships of the signal model stated in (2)-(4) and (26), the MSE becomes

$$J = P_{\mathbf{p}_0}\left\{\left[1+(k^2+1)\frac{1-\gamma}{2\gamma}\right]w_{\mathrm{P}0,0}{}^2 + \left[k^2+(k^2+1)\frac{1-\gamma}{2\gamma}\right]w_{\mathrm{P}0,1}{}^2 - 2w_{\mathrm{P}0,0} - 2k w_{\mathrm{P}0,1} + 2k w_{\mathrm{P}0,0} w_{\mathrm{P}0,1} + 1\right\}. \tag{27}$$

Hence, the weights can be easily obtained by taking the gradients of J with respect to $w_{\mathrm{P}0,0}, w_{\mathrm{P}0,1}$ and equating their results to zero. The weights of the primary component extracted by LS are found to be $w_{\mathrm{P}0,0} = \frac{2\gamma}{1+\gamma}\frac{1}{1+k^2}$, $w_{\mathrm{P}0,1} = \frac{2\gamma}{1+\gamma}\frac{k}{1+k^2}$. Similarly, the weights for the remaining components can also be derived. The extracted primary and ambient components using LS are thus expressed as

$$\begin{aligned}\hat{\mathbf{p}}_{\mathrm{LS},\,0} &= \frac{2\gamma}{1+\gamma}\frac{1}{1+k^2}\left(\mathbf{x}_0 + k\mathbf{x}_1\right),\\ \hat{\mathbf{p}}_{\mathrm{LS},\,1} &= \frac{2\gamma}{1+\gamma}\frac{k}{1+k^2}\left(\mathbf{x}_0 + k\mathbf{x}_1\right),\end{aligned} \tag{28}$$

$$\begin{aligned}\hat{\mathbf{a}}_{\mathrm{LS},0} &= \frac{1+k^2+\left(k^2-1\right)\gamma}{1+\gamma}\frac{1}{1+k^2}\mathbf{x}_0 - \frac{2\gamma}{1+\gamma}\frac{k}{1+k^2}\mathbf{x}_1,\\ \hat{\mathbf{a}}_{\mathrm{LS},1} &= -\frac{2\gamma}{1+\gamma}\frac{k}{1+k^2}\mathbf{x}_0 + \frac{1+k^2+\left(1-k^2\right)\gamma}{1+\gamma}\frac{1}{1+k^2}\mathbf{x}_1.\end{aligned} \tag{29}$$

From (28)-(29), we observe that the weights for the extracted primary and ambient components are not only dependent on *k*, but also related to *γ*. As compared with PCA, the panning relationship of *k* between the extracted primary components in the two channels still holds, but no explicit panning is found in the extracted ambient components using LS.

C. Primary-Ambient Extraction Using Minimum Leakage Least Squares

As discussed in Section III, three types of error may be found in the extracted components, namely, the distortion, interference, and leakage. The leakage is the most undesirable among the three, and priority should be given to the minimization of the leakage in the extraction process. We therefore propose MLLS, which minimizes the extraction error with the constraint that the leakage is minimum in the extracted components. The amount of leakage power in the extracted primary or ambient component can be quantified by the leakage-to-extracted-signal ratio (LeSR), which is given as

$$\mathrm{LeSR}_{\mathrm{P}} = P_{Leak_{\mathrm{P}}} \big/ P_{\hat{\mathbf{p}}_0}, \quad \mathrm{LeSR}_{\mathrm{A}} = P_{Leak_{\mathrm{A}}} \big/ P_{\hat{\mathbf{a}}_0}. \tag{30}$$

Minimum leakage in the extracted components is achieved by minimizing LeSR. For the extracted primary component, the leakage comes from the ambient components. Using (15) and (30), the LeSR$_{\mathrm{P}}$ is computed as:

$$\mathrm{LeSR}_{\mathrm{P}} = \frac{\left(w_{\mathrm{P}0,0}{}^2 + w_{\mathrm{P}0,1}{}^2\right)P_{\mathbf{a}_0}}{\left(w_{\mathrm{P}0,0} + k w_{\mathrm{P}0,1}\right)^2 P_{\mathbf{p}_0} + \left(w_{\mathrm{P}0,0}{}^2 + w_{\mathrm{P}0,1}{}^2\right)P_{\mathbf{a}_0}}. \tag{31}$$

Minimizing LeSR_P with respect to $w_{\mathrm{P}0,0}, w_{\mathrm{P}0,1}$, we have

$$w_{\mathrm{P}0,1} = k w_{\mathrm{P}0,0}. \tag{32}$$

Next, we substitute (32) into the extraction error given by (15), and the extraction error becomes

$$\boldsymbol{\varepsilon}_{\mathrm{P}} = \left[\left(1+k^2\right) w_{\mathrm{P}0,0} - 1\right] \mathbf{p}_0 + w_{\mathrm{P}0,0}\mathbf{a}_0 + k w_{\mathrm{P}0,0}\mathbf{a}_1. \tag{33}$$

Based on (12) and (33), the ESR$_{\mathrm{P}}$ is expressed as

$$\mathrm{ESR}_{\mathrm{P}} = \frac{\left[\left(1+k^2\right) w_{\mathrm{P}0,0} - 1\right]^2 P_{\mathbf{p}_0} + \left(w_{\mathrm{P}0,0}{}^2 + k^2 w_{\mathrm{P}0,0}{}^2\right) P_{\mathbf{a}_0}}{P_{\mathbf{p}_0}}. \tag{34}$$

By minimizing ESR$_{\mathrm{P}}$, we arrive at $w_{\mathrm{P}0,0} = \frac{2\gamma}{1+\gamma}\frac{1}{1+k^2}$, and $w_{\mathrm{P}0,1} = \frac{2\gamma}{1+\gamma}\frac{k}{1+k^2}$. Finally, we can express the primary component in channel 0 extracted by MLLS as

$$\hat{\mathbf{p}}_{\mathrm{MLLS},0} = \frac{2\gamma}{1+\gamma}\frac{1}{1+k^2}\left(\mathbf{x}_0 + k\mathbf{x}_1\right), \tag{35}$$

The remaining components extracted by MLLS can be obtained similarly, and are found to be

$$\hat{\mathbf{p}}_{\mathrm{MLLS},1} = \frac{2\gamma}{1+\gamma}\frac{k}{1+k^2}\left(\mathbf{x}_0 + k\mathbf{x}_1\right), \tag{36}$$

$$\begin{aligned}\hat{\mathbf{a}}_{\mathrm{MLLS},0} &= \frac{k}{1+k^2}\left(k\mathbf{x}_0 - \mathbf{x}_1\right),\\ \hat{\mathbf{a}}_{\mathrm{MLLS},1} &= -\frac{1}{1+k^2}\left(k\mathbf{x}_0 - \mathbf{x}_1\right).\end{aligned} \tag{37}$$

D. Primary-Ambient Extraction Using Minimum Distortion Least Squares

Inspired by the popular minimum variance distortionless response (MVDR) filter [45], we propose the minimum distortion least squares in PAE by minimizing the extraction error ESR, with the constraint that the extracted component is distortionless. Mathematically, we can express the objective function of MDLS as $\min_{\mathbf{w}} \text{ESR s.t. DSR} = 0.$ Similar to the steps in MLLS, the solution for each extracted component is easily found to be

$$\begin{aligned}\hat{\mathbf{p}}_{\text{MDLS},0} &= \frac{1}{1+k^2}\left(\mathbf{x}_0 + k\mathbf{x}_1\right),\\ \hat{\mathbf{p}}_{\text{MDLS},1} &= \frac{k}{1+k^2}\left(\mathbf{x}_0 + k\mathbf{x}_1\right).\end{aligned}\tag{38}$$

$$\begin{aligned}\hat{\mathbf{a}}_{\text{MDLS},0} &= \mathbf{x}_0 - \frac{2k\gamma}{\left(k^2-1\right)\gamma + k^2 + 1}\mathbf{x}_1,\\ \hat{\mathbf{a}}_{\text{MDLS},1} &= -\frac{2k\gamma}{\left(1-k^2\right)\gamma + k^2 + 1}\mathbf{x}_0 + \mathbf{x}_1.\end{aligned}\tag{39}$$

E. Comparison among PCA, LS, MLLS, and MDLS in PAE

In this subsection, we compare the relationships and differences, as well as the performance among the four linear estimation based PAE approaches. The key minimization criteria and relationships of these approaches are illustrated in Fig. 2. Based on the linear estimation framework, PCA minimizes the correlation between the primary and ambient components, whereas LS, MLLS, and MDLS aim to minimize the extraction error, leakage, and distortion, respectively, for both the primary and ambient components. Some interesting relationships can be found for the primary components extracted using these approaches. From (22) and (38), we find that $\hat{\mathbf{p}}_{\text{PCA},i} = \hat{\mathbf{p}}_{\text{MDLS},i}, \forall i \in \{0,1\}.$ This equivalence implies that PCA extracts the primary component with minimum distortion, even though PCA does not explicitly specify this constraint as found in MDLS. From (28) and (35)-(36), we observe that $\hat{\mathbf{p}}_{\text{LS},i} = \hat{\mathbf{p}}_{\text{MLLS},i}.$ This equivalence implies that LS extracts the primary component with minimum leakage, even though LS does not explicitly specify this constraint as found in MLLS. There is an amplitude difference between the primary components extracted by MLLS and by MDLS, i.e.,

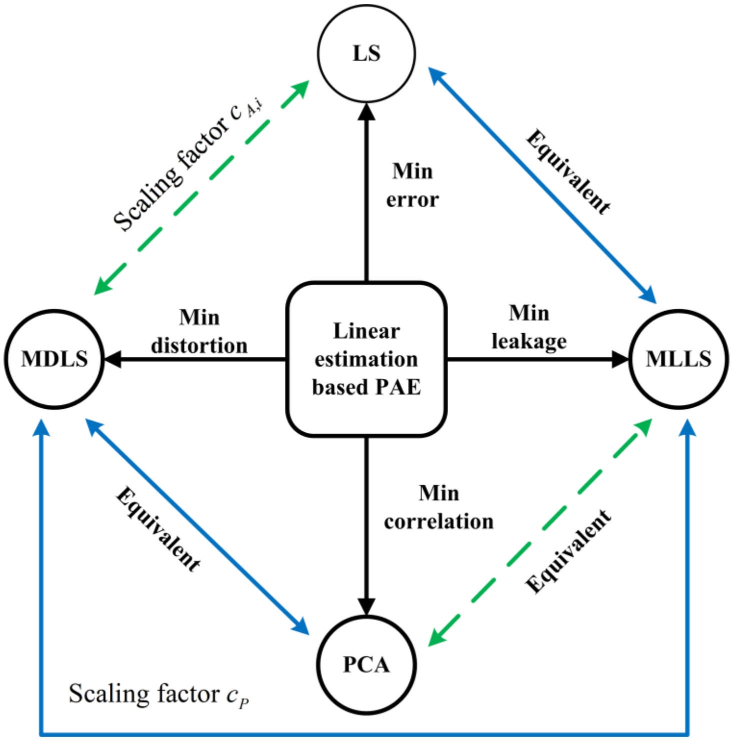


Fig. 2. Objectives and relationships of four linear estimation based PAE approaches. Blue solid lines represent the relationships in the primary component, and green dotted lines represent the relationships in the ambient component.

$$\hat{\mathbf{p}}_{\mathrm{MLLS},i} = c_{\mathrm{P}} \hat{\mathbf{p}}_{\mathrm{MDLS},i}, \tag{40}$$

where the scaling factor $c_{\mathrm{P}} = 2\gamma/(1+\gamma)$. Since $\gamma \in [0,1]$, $c_{\mathrm{P}} \leq 1$, it is clear that the primary component extracted by MLLS has lower power than the primary component extracted by MDLS for all $\gamma \neq 1$.

Similarly, we noted a few interesting relationships for the extracted ambient component. Based on (23) and (37), it is interesting to find that $\hat{\mathbf{a}}_{\mathrm{PCA},i} = \hat{\mathbf{a}}_{\mathrm{MLLS},i}$. This equivalence implies that PCA extracts the ambient component with minimum leakage, even though PCA does not explicitly specify this constraint as found in MLLS. From (29) and (39), there is also an amplitude difference between the ambient components extracted by MDLS and LS, which is given by

Table I

Results of measures for PCA, LS, minimum leakage LS, and minimum distortion LS in PAE.

Measures		**Primary component**		**Ambient component**		
		MDLS/ PCA	**MLLS/LS**	**MLLS/PCA**	**LS**	**MDLS**
Group 1: Extraction Accuracy	Error-to-signal ratio, ESR	$\frac{1-\gamma}{2\gamma}$	$\frac{1-\gamma}{1+\gamma}$	$\frac{1}{1+k^2}$	$\frac{1}{1+k^2}\frac{2\gamma}{1+\gamma}$	$\frac{2\gamma}{\left(k^2-1\right)\gamma+k^2+1}$
	Leakage-to-signal ratio, LSR	$\frac{1-\gamma}{2\gamma}$	$\frac{1-\gamma}{2\gamma}\left(\frac{2\gamma}{1+\gamma}\right)^2$	0	$\frac{1}{1+k^2}\frac{2\gamma(1-\gamma)}{(1+\gamma)^2}$	$\frac{\left(1+k^2\right)(1-\gamma)2\gamma}{\left[\left(1+k^2\right)(1+\gamma)-2\gamma\right]^2}$
	Distortion-to-signal ratio, DSR	0	$\left(\frac{1-\gamma}{1+\gamma}\right)^2$	$\left(\frac{1}{1+k^2}\right)^2$	$\left(\frac{1}{1+k^2}\frac{2\gamma}{1+\gamma}\right)^2$	0
	Interference-to-signal ratio, ISR	0		$\left(\frac{k}{1+k^2}\right)^2$	$\left(\frac{k}{1+k^2}\frac{2\gamma}{1+\gamma}\right)^2$	$\left[\frac{2k\gamma}{\left(1+k^2\right)(1+\gamma)-2\gamma}\right]^2$
Group 2: Spatial Accuracy	ICC(ICTD)	1(0)		1	$\frac{2k\gamma}{\sqrt{\left(1+k^2\right)^2-\left(1-k^2\right)^2\gamma^2}}$	
	ICLD	k^2		$\frac{1}{k^2}$	$\frac{1}{k^2}\frac{1+\gamma+k^2(1-\gamma)}{1+\gamma+\frac{1}{k^2}(1-\gamma)}$	$\frac{1}{k^2}\frac{1-\gamma+k^2(1+\gamma)}{1-\gamma+\frac{1}{k^2}(1+\gamma)}$

γ denotes the primary power ratio PPR, and k represents the primary panning factor PPF.

$$\hat{\mathbf{a}}_{\mathrm{LS},i}=c_{\mathrm{A},i}\hat{\mathbf{a}}_{\mathrm{MDLS},i}, \tag{41}$$

where $c_{\mathrm{A},i}=\frac{1+k^2+(-1)^i\left(k^2-1\right)\gamma}{\left(1+k^2\right)(1+\gamma)}$. As compared to (40), the scaling factor in the extracted ambient components differs from channel 0 to channel 1.

Next, we present a comparative analysis on the performance of these four PAE approaches. Here, we summarize the results of the performance measures obtained with channel 0 in Table I. Due to the symmetry in the stereo signal model, the measures for channel 1 can be obtained by replacing k in the results in Table I with its reciprocal. From Table I, it is clear that the two groups of measures are highly dependent on γ and/or k.

For the primary extraction, we have the following observations of MDLS (or PCA) and MLLS (or LS) based on the measures in Table I. In Group 1, lower ESR and LSR of the extracted primary component are observed in MLLS as compared to MDLS. The distortion measure DSR = 0 indicates that primary component extracted using MDLS (or PCA) is free of distortion, whereas the distortion in MLLS (or LS) increases as γ decreases. Hence, MLLS (or LS) extracts primary component with minimum leakage and error at the expense of introducing some distortion in the extracted primary component. All four approaches extract primary component without interference. According to the spatial cues (ICC, ICTD, and ICLD) of the primary component in Group 2, all four approaches are capable of preserving the correct spatial information in the extracted primary component.

For the ambient extraction, we have the following observations of MLLS (or PCA), LS, and MDLS based on the measures in Table I. In Group 1, we observe that LS has the lowest ESR. The measure LSR = 0 found in MLLS indicates that no primary components are leaked into the extracted ambient component. By contrast, a certain amount of primary leakage is found in ambient component extracted using LS or MDLS. As for DSR, only MDLS extracts the ambient component without distortion. The overall best performance on the ambient extraction is achieved using LS based on the measures of diffuseness in Group 2, but none of the approaches is able to extract an uncorrelated and balanced ambient component. Therefore, some post-processing techniques such as decorrelation [46] and post-scaling [25] should be used to enhance the ambient extraction.

F. Primary-Ambient Extraction Using Adjustable Least Squares

In this subsection, we propose the adjustable least squares, which is designed to achieve an adjustable performance in terms of extraction error and distortion, as well as producing minimum leakage in the extracted primary and ambient components. Similar to (32), by minimizing the leakage LeSR in the extracted primary and ambient components, we have $\left[w_{\mathrm{P}0,1}, w_{\mathrm{P}1,1}\right] = k\left[w_{\mathrm{P}0,0}, w_{\mathrm{P}1,0}\right]$, and $\left[w_{\mathrm{A}0,1}, w_{\mathrm{A}1,1}\right] = -k^{-1}\left[w_{\mathrm{A}0,0}, w_{\mathrm{A}1,0}\right]$, respectively. To achieve the adjustable performance in terms of

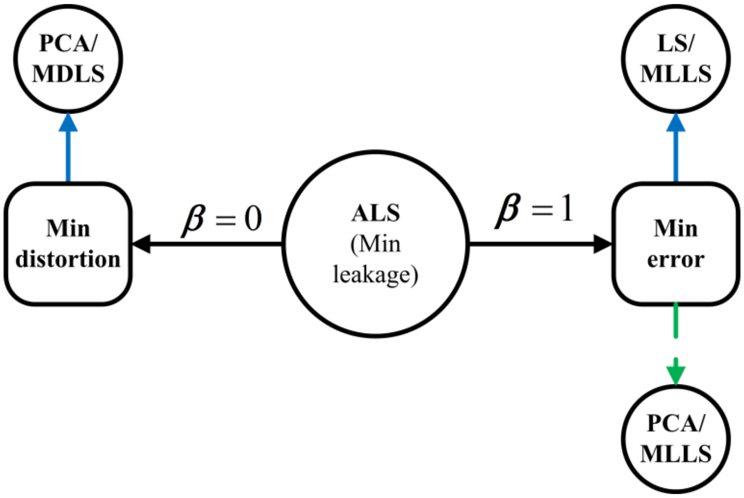


Fig. 3. Characteristics and relationships of adjustable least squares. Blue solid lines represent the relationships in the primary component, and green dotted lines represent the relationships in the ambient component.

extraction error and distortion, we introduce the adjustable factor β where $0 \le \beta \le 1$. By letting β = 0, and β = 1, we can achieve the minimum distortion and extraction error, respectively. Based on our analysis of the four PAE approaches, the weights in ALS are obtained as

$$\begin{bmatrix} w_{\mathrm{P}0,0} & w_{\mathrm{P}0,1} \\ w_{\mathrm{P}1,0} & w_{\mathrm{P}1,1} \end{bmatrix} = \frac{1}{1+k^2}\left(1-\beta\frac{1-\gamma}{1+\gamma}\right)\begin{bmatrix} 1 & k \\ k & k^2 \end{bmatrix}, \tag{42}$$

$$\begin{bmatrix} w_{\mathrm{A}0,0} & w_{\mathrm{A}0,1} \\ w_{\mathrm{A}1,0} & w_{\mathrm{A}1,1} \end{bmatrix} = \begin{bmatrix} 1-\beta\frac{1}{1+k^2} & -\frac{1}{k}\left(1-\beta\frac{1}{1+k^2}\right) \\ -k\left(1-\beta\frac{k^2}{1+k^2}\right) & 1-\beta\frac{k^2}{1+k^2} \end{bmatrix}. \tag{43}$$

Next, the three key performance measures for PAE using ALS are expressed as

$$\begin{aligned} &\mathrm{ESR}_{\mathrm{P}} = \frac{1-\gamma}{2\gamma} + \beta(\beta-2)\frac{(1-\gamma)^2}{2\gamma(1+\gamma)},\ \ \mathrm{DSR}_{\mathrm{P}} = \beta^2\left(\frac{1-\gamma}{1+\gamma}\right)^2,\ \ \mathrm{LeSR}_{\mathrm{P}} = \frac{1-\gamma}{1+\gamma}, \\ &\mathrm{ESR}_{\mathrm{A}} = \frac{1}{k^2} + \beta(\beta-2)\frac{1}{k^2\left(k^2+1\right)},\ \ \mathrm{DSR}_{\mathrm{A}} = \beta^2\left(\frac{1}{1+k^2}\right)^2,\ \mathrm{LeSR}_{\mathrm{A}} = 0. \end{aligned} \tag{44}$$

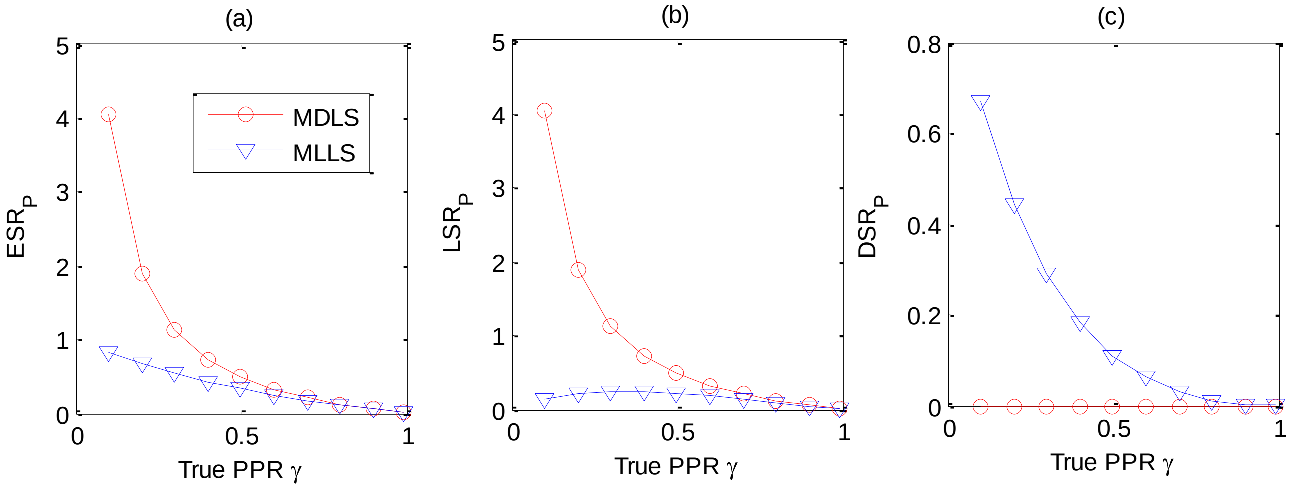


Fig. 4. Comparison of MDLS (or PCA) and MLLS (or LS) in primary extraction, (a) error-to-signal ratio ESR; (b) leakage-to-signal ratio LSR, (c) distortion-to-signal ratio DSR. Legend in (a) applies to all plots.

From the above measures, it can be inferred that the extraction error ESR decreases and the distortion DSR increases gradually as β increases, whereas the measure for leakage LeSR remains constant and small. Since the adjustable factor β = 0 and β = 1 led to minimum distortion and extraction error, respectively, other values of β between 0 and 1 yield an adjustable performance in terms of extraction error and distortion. For example, ALS with β = 0.5 produces 75% reduction of extraction error and distortion in PAE. The characteristics of ALS and its relationships with other PAE approaches are illustrated in Fig. 3. By adjusting the value of β, ALS can achieve the performance of the previously discussed PAE approaches. Specifically, in primary extraction, ALS with β = 0 is equivalent to MDLS (or PCA), whereas ALS with β = 1 is equivalent to MLLS (or LS). In ambient extraction, ALS can be linked with MLLS (or PCA) by letting β = 1.

V. EXPERIMENTAL RESULTS AND DISCUSSIONS

Since our focus in this paper is to compare different linear estimation based PAE approaches rather than the subband decomposition of the stereo signal, we shall consider only one primary component in the stereo signal. Hence, no subband decomposition of the stereo signal is considered in our simulations. A speech signal is selected as the primary component and uncorrelated white Gaussian noise with equal variance in two channels is synthesized as the ambient component in our simulations. To simulate the source panned to channel 1, the primary component is scaled by $k = 5$. Subsequently, the stereo signals are synthesized by linearly mixing the primary and ambient components using different values of primary power ratio PPR, ranging from zero to one. The performance of these PAE approaches is then evaluated using the performance measures introduced in Section III. Based on our simulations, we provide some recommendations for the applications using these PAE approaches.

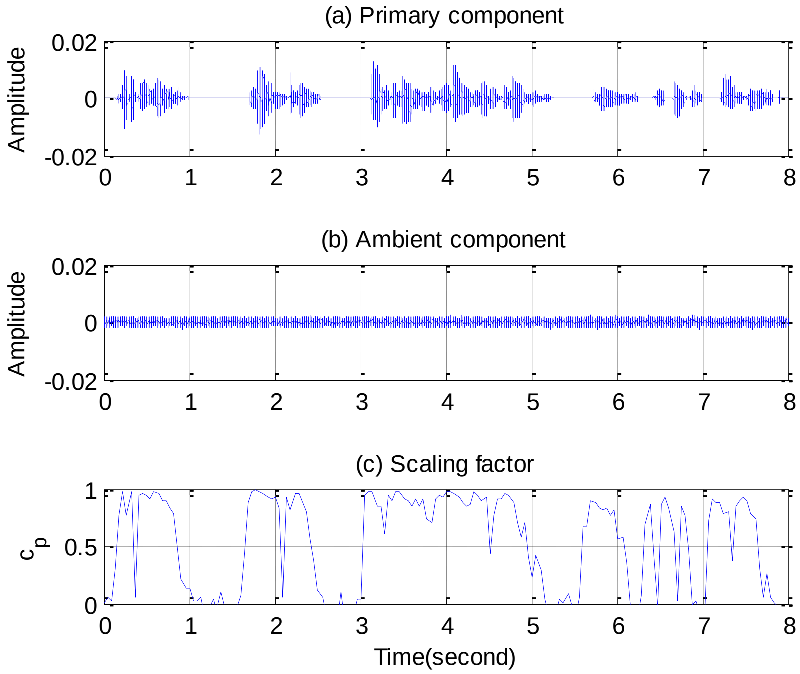


Fig. 5. Scaling difference between the primary components extracted using MLLS and MDLS.

A. Comparison of PAE Using PCA, LS, MLLS and MDLS

The simulation results of PAE using PCA, LS, MLLS, and MDLS are shown in Figs. 4-7. Recall that the extraction performance of the primary component is identical for PCA and LS with MDLS and MLLS, respectively, we shall discuss the primary extraction for MLLS and MDLS only in this subsection. The extraction accuracy of the extracted primary components using MLLS and MDLS (same for the two channels) is shown in Fig. 4. Several observations from Fig. 4 are as follows. The extraction error given by ESR_P reduces gradually as γ increases. The ESR_P and LSR_P for MLLS are relatively lower than those in MDLS, which indicates that MLLS is superior to MDLS in extracting the primary component in terms of the extraction error and leakage. However, the distortion of extracted primary component using MLLS increases as γ decreases, while no distortion is found with MDLS.

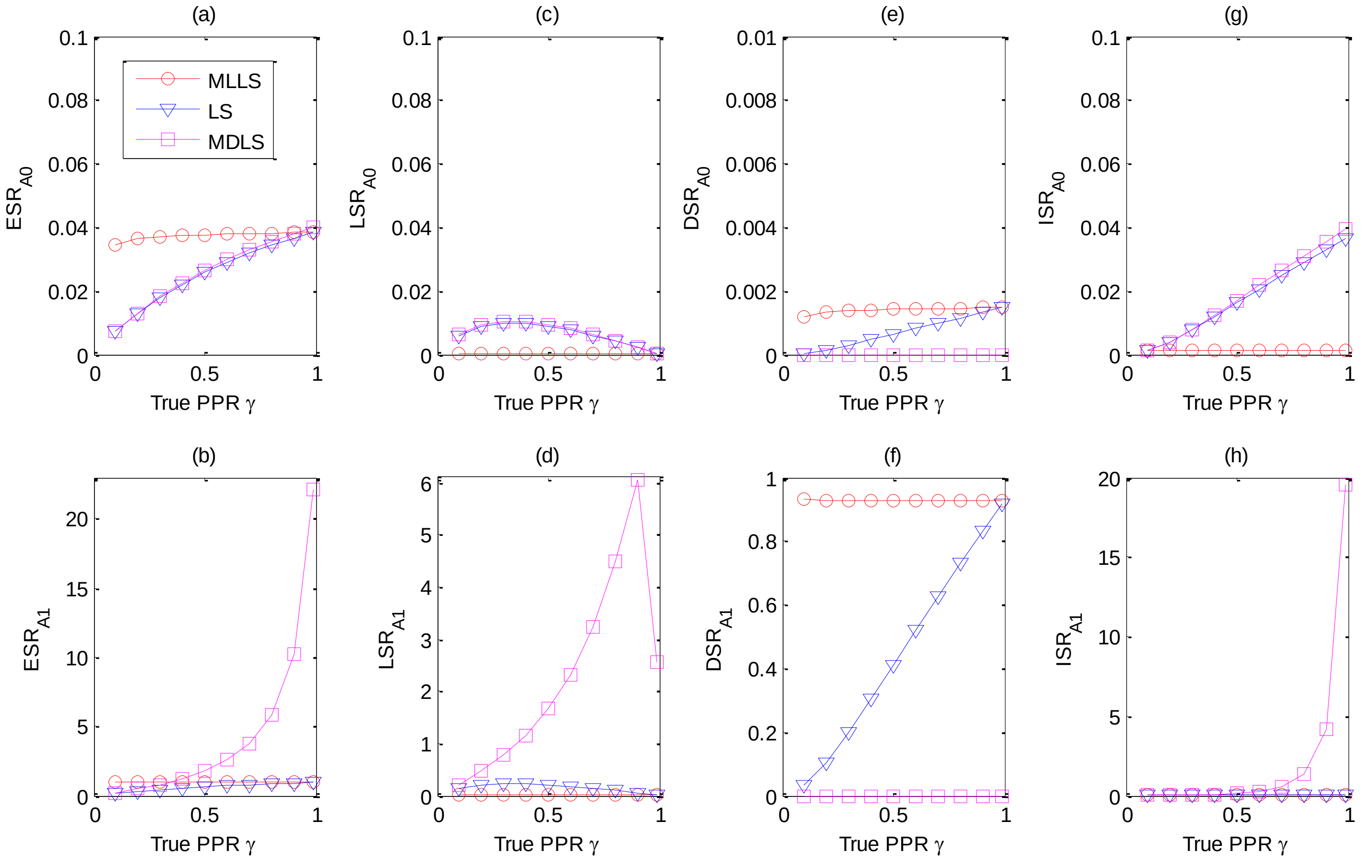


Fig. 6. Comparison of ambient extraction with MLLS (or PCA), LS and MDLS for channel 0 (top row) and channel 1 (bottom row). (a)-(b) error-to-signal ratio ESR; (c)-(d) leakage-to-signal ratio LSR; (e)-(f) distortion-to-signal ratio DSR; (g)-(h) interference-to-signal ratio ISR. Legend in (a) applies to all plots.

The difference in the performance for the extracted primary component between MLLS and MDLS is caused by the scaling difference, as expressed in (40). This scaling factor depends solely on PPR, which is determined by the power difference between true primary and ambient components in each frame. In the case of stationary primary and ambient components, the scaling factor is almost constant and leading to similar performance between MLLS and MDLS. However, there is a noticeable difference in the primary components extracted using MLLS and MDLS when the primary component is non-stationary. An example to illustrate the variation of the scaling factor is shown in Fig. 5. It is observed that the scaling factor is

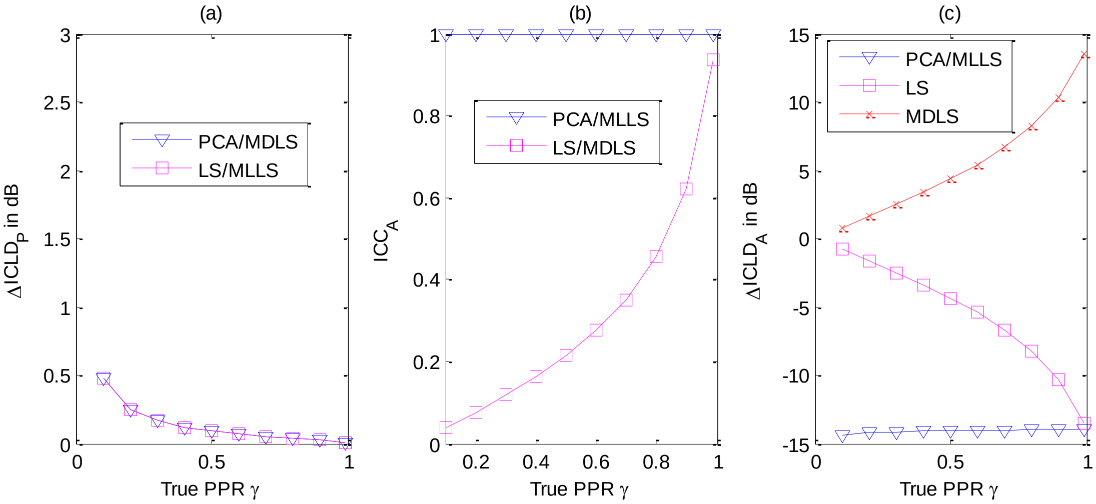


Fig. 7. Comparison of spatial accuracy in PCA, LS, MLLS, and MDLS. (a) ICLD estimation error in the extracted primary component; (b) ICC of the extracted ambient component; and (c) ICLD estimation error in the extracted ambient component.

fluctuating according to the power difference between primary and ambient components. The scaling factor rises closer to one when the primary component power is comparably stronger than the ambient component power, and the scaling factor drops to zero when the primary component becomes relatively weak compared to the ambient component. This example reveals that MLLS and MDLS behave similarly when primary component is dominant and only MLLS can extract weak primary component at the ambient-dominant periods of the signal. As a result, MLLS has lower ESR$_P$ but the extracted primary component may possess some discontinuity and more distortion.

Ambient extraction using PCA, LS, MLLS and MDLS is illustrated in Fig. 6. Unlike the primary extraction, the performance of ambient extraction has significant variation between the two channels. Due to the weaker primary component in channel 0, the performance of ambient extraction in channel 0 is better than that in channel 1 as shown in our simulations. Nevertheless, some common characteristics in the performance of ambient extraction in the two channels are observed. We found that LS has the lowest extraction error, whereas MLLS (or PCA), and MDLS can completely remove the leakage and distortion,

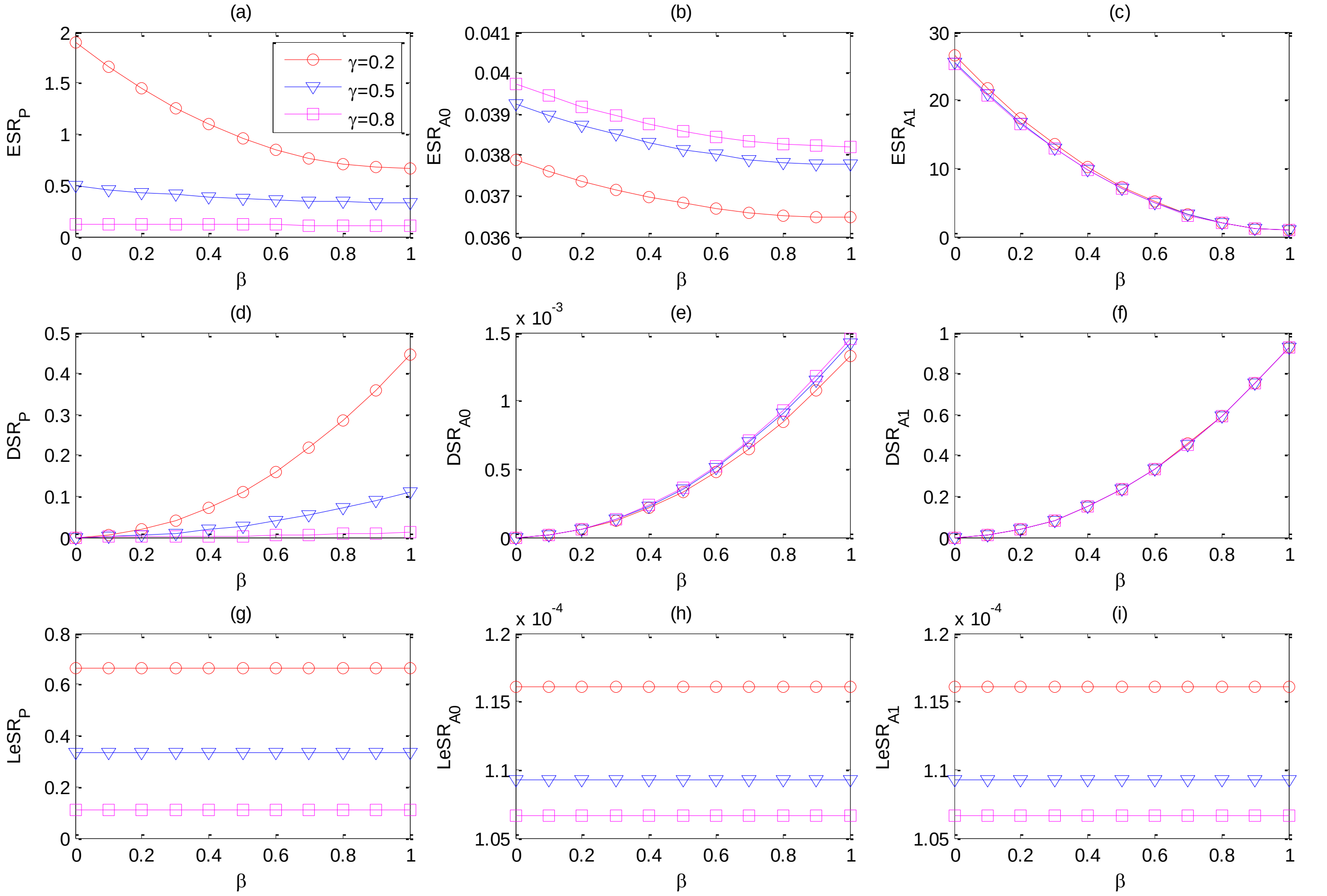


Fig. 8. Measures for ALS with different values of adjustable factor β, error-to-signal ratio ESR (top row), distortion-to-signal ratio DSR (middle row), and leakage-to-extracted-signal ratio LeSR (bottom row), for the primary component (left column), the ambient component in channel 0 (middle column), and the ambient component in channel 1 (right column). Legend in (a) applies to all plots. Three lines in each plot represent different values of PPR γ.

respectively. However, MDLS extracts the ambient component in channel 1 with much higher extraction error, leakage, and interference than the other PAE approaches.

Finally, we examine the spatial accuracy of the extracted primary and ambient components, as shown in Fig. 7. Since the extracted primary components are all scaled by k between the two channels, the ICC and ICTD of the primary components are exactly the same as the true values, and the ICLD$_P$ is also very close to its true value, as shown in Fig. 7(a). However, from the results of ICC$_A$ and ICLD$_A$ shown in Fig. 7(b) and 7(c), respectively, we found that none of these approaches are able to extract uncorrelated and balanced ambient components.

B. Performance of ALS in PAE

The performance of PAE using ALS is shown in Fig. 8. The measures for extraction error, distortion, and leakage are examined with respect to the adjustable factor β. These measures for the primary components for both channels are presented in the plots in the left column. The results of the measures for ambient extraction for the channels 0 and 1 are presented in the plots in the middle and right columns, respectively. From the plots of the top and middle rows, we observed that larger values of β lead to lower extraction error (as shown by ESR) but higher distortion (as shown by DSR). Nevertheless, the leakage as quantified by LeSR remains at a very low level for all values of β, as shown in the plots in the bottom row. These observations verified that the adjustable performance in terms of extraction error and distortion using ALS is achieved by adjusting β.

C. General Guidelines in Selecting PAE Approaches

Generally, the selection of the PAE approach depends on the post-processing techniques and playback systems which are associated with the specific audio application, as well as the audio content and user preferences. Several guidelines on the applications of these PAE approaches can be drawn from our analysis and discussions. We summarize the strengths, weaknesses, and our recommendations of these PAE approaches in Table II. In applications like spatial audio coding and interactive audio in gaming, where the primary component is usually more important than the ambient component, PCA would be a better choice. In the case where both the primary and ambient components are extracted, processed, and finally mixed

Table II

Strengths, weaknesses, and recommendations of different PAE approaches

Approaches	Strengths	Weaknesses	Recommendations
PCA	• No distortion in the extracted primary component; • No primary leakage in the extracted ambient component; • Primary and ambient components are uncorrelated;	• Ambient component severely panned;	Spatial audio coding and interactive audio in gaming, where the primary component is more important than the ambient component.
LS	• Minimum MSE in the extracted primary and ambient components;	• Severe primary leakage in the extracted ambient component;	Applications in which both the primary and ambient components are extracted, processed, and finally mixed together.
MLLS	• Minimum leakage in the extracted primary and ambient components; • Primary and ambient components are uncorrelated;	• Ambient component severely panned;	Spatial audio enhancement systems, and applications in which different rendering or playback techniques are employed on the extracted primary and ambient components.
MDLS	• No distortion in the extracted primary and ambient components;	• Severe interference and primary leakage in the extracted ambient component;	High-fidelity applications in which timbre is of high importance.
ALS	• Performance adjustable;	• Need to adjust the value of the adjustable factor;	For applications without explicit requirements.

together, the extraction error becomes more critical and hence LS is recommended. In some spatial audio enhancement systems, where the extracted primary or ambient component is added back to the original signal to emphasize the extracted component, accurate extraction of the primary or ambient component becomes the key consideration. For such systems, MLLS is preferred as the leakage becomes the most important consideration. MLLS is also recommended when different rendering and playback techniques are employed on the extracted primary and ambient components. MDLS is more suitable for high-fidelity applications, where timbre is of high importance, such as in musical application. When there is no explicit requirement, ALS can be employed by setting the proper adjustable factor.

VI. CONCLUSIONS

In this paper, we revisited the problem of primary-ambient extraction (PAE) of stereo signals using linear estimation based approaches. Based on the stereo signal model, we formulated the PAE problem as a problem to determine the weighting matrix under our linear estimation framework. Under this framework, we introduced two groups of performance measures and derived the solutions for two existing approaches, namely, principal component analysis (PCA), and least squares (LS). Based on the objectives of minimum leakage, minimum distortion, and adjustable performance, we proposed three additional approaches, namely, minimum leakage least squares (MLLS), minimum distortion least squares (MDLS), and adjustable least squares (ALS). The relationships and differences of these PAE approaches are extensively studied. For primary extraction, PCA was found to be equivalent to MDLS in terms of minimum distortion; and LS is equivalent to MLLS in terms of minimum extraction error and leakage. The difference between extracted primary components using MDLS and MLLS is found to be a scaling factor, which is solely related to primary power ratio (PPR). All the discussed PAE approaches perform well for primary extraction but perform poorly in extracting ambient component when PPR is high. In ambient extraction, MLLS (or PCA), LS, and MDLS minimize the leakage, extraction error, and distortion, respectively. Adjustable LS offers an adjustable performance in terms of extraction error and distortion with the constraint of minimum leakage. Based on our discussions in this paper, these PAE approaches are suggested in different spatial audio applications.

Appendix

Derivation of (22) from (21) in Section IV.A

We show the derivations for the extracted primary component in channel 0. From (8) and (19), we can find

$$\frac{\lambda_{\mathrm{P}} - r_{00}}{r_{01}} = k. \tag{45}$$

From (5)-(7), we have

$$r_{11} = \frac{k^2 - 1}{k} r_{01} + r_{00}. \tag{46}$$

Based on (45), we can rewrite (20) as

$$\mathbf{u}_{\mathrm{P}} = r_{01}\left(\mathbf{x}_0 + k\mathbf{x}_1\right). \tag{47}$$

Substitute (47) into (21),

$$\begin{aligned}
\hat{\mathbf{p}}_0 &= \frac{\mathbf{u}_{\mathrm{P}}{}^H \mathbf{x}_0}{\mathbf{u}_{\mathrm{P}}{}^H \mathbf{u}_{\mathrm{P}}} \mathbf{u}_{\mathrm{P}} \\
&= \frac{r_{01}\left(\mathbf{x}_0 + k\mathbf{x}_1\right)^H \mathbf{x}_0}{r_{01}\left(\mathbf{x}_0 + k\mathbf{x}_1\right)^H r_{01}\left(\mathbf{x}_0 + k\mathbf{x}_1\right)} r_{01}\left(\mathbf{x}_0 + k\mathbf{x}_1\right). \\
&= \frac{r_{01}\left(r_{00} + kr_{01}\right)}{r_{01}{}^2\left(r_{00} + k^2 r_{11} + 2kr_{01}\right)} r_{01}\left(\mathbf{x}_0 + k\mathbf{x}_1\right) \\
&= \frac{r_{00} + kr_{01}}{r_{00} + k^2 r_{11} + 2kr_{01}}\left(\mathbf{x}_0 + k\mathbf{x}_1\right)
\end{aligned} \tag{48}$$

Substitute (46) into (48),

$$\begin{aligned}
\hat{\mathbf{p}}_0 &= \frac{r_{00} + kr_{01}}{r_{00} + k^2\left(\dfrac{k^2 - 1}{k} r_{01} + r_{00}\right) + 2kr_{01}}\left(\mathbf{x}_0 + k\mathbf{x}_1\right) \\
&= \frac{r_{00} + kr_{01}}{\left(1 + k^2\right) r_{00} + k\left(k^2 + 1\right) r_{01}}\left(\mathbf{x}_0 + k\mathbf{x}_1\right) \\
&= \frac{1}{1 + k^2}\left(\mathbf{x}_0 + k\mathbf{x}_1\right).
\end{aligned} \tag{49}$$

Thus, we obtain the simplified expression of the extracted primary component in channel 0, as shown in (22). The primary component in channel 1 and the ambient components can also be derived in the same way.

ACKNOWNLEDGEMENT

The authors wish to thank the reviewers for their constructive comments and suggestions.

REFERENCES

[1] 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, Jul./Aug. 2004.

[2] M. M. Goodwin and J. M. Jot, "Primary-ambient signal decomposition and vector-based localization for spatial audio coding and enhancement," *in Proc. ICASSP*, Hawaii, 2007, pp. 9-12.

[3] F. Menzer and C. Faller, "Stereo-to-binaural conversion using interaural coherence matching", in *Proc. 128th Audio Eng. Soc. Conv.*, London, UK, 2010.

[4] T. Holman, *Surround sound up and running 2nd ed*., MA: Focal Press, 2008.

[5] F. Rumsey, *Spatial Audio*. Oxford, UK: Focal Press, 2001.

[6] J. Breebaart and C. Faller, *Spatial audio processing: MPEG surround and other applications*. Chichester, UK: John Wiley & Sons, 2007.

[7] J. Breebaart and E. Schuijers, "Phantom materialization: a novel method to enhance stereo audio reproduction on headphones," *IEEE Trans. Audio, Speech, Lang. Process.*, vol.16, no. 8, pp. 1503-1511, Nov. 2008.

[8] M. M. Goodwin and J. M. Jot, "Binaural 3-D audio rendering based on spatial audio scene coding," in *Proc. 123rd Audio Eng. Soc. Conv.*, New York, 2007.

[9] C. Faller and J. Breebaart, "Binaural reproduction of stereo signals using upmixing and diffuse rendering," in *Proc. 131th Audio Eng. Soc. Conv.*, New York, 2011.

[10] V. Pulkki, "Spatial sound reproduction with directional audio coding," *J. Audio Eng. Soc.*, vol. 55, no. 6, pp. 503-516, Jun. 2007.

[11] M. M. Goodwin and J. M. Jot, "Spatial audio scene coding," in *Proc. 125th Audio Eng. Soc. Conv.*, San Francisco, 2008.

[12] M. R. Bai and G. Y. Shih, "Upmixing and downmixing two-channel stereo audio for consumer electronics," *IEEE Trans. Consumer Electron.*, vol. 53, no. 3, pp. 1011-1019, Aug. 2007.

[13] S. Y. Park, S. Lee, and D. Youn, "Robust representation of spatial sound in stereo-to-multichannel upmix," in *Proc. 128th Audio Eng. Soc. Conv.*, London, UK, 2010.

[14] W. S. Gan, E. L. Tan, and S. M. Kuo, "Audio projection: directional sound and its application in immersive communication," *IEEE Signal Process. Mag.*, vol. 28, no. 1, pp. 43-57, Jan. 2011.

[15] E. L. Tan, and W. S. Gan, "Reproduction of immersive sound using directional and conventional loudspeakers," *J. Acoust. Soc. Am.*, vol. 131, no. 4, pp. 3215-3215, Apr. 2012.

[16] E. L. Tan, W. S. Gan, and C. H. Chen, "Spatial sound reproduction using conventional and parametric loudspeakers," in *Proc. APSIPA ASC*, Hollywood, CA, 2012.

[17] F. Baumgarte and C. Faller, "Binaural cue coding-part I: psychoacoustic fundamental and design principles," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 6, pp. 509–519, Nov. 2003.

[18] J. Breebaart, G. Hotho, J. Koppens, E. Schuijers, W. Oomen and S. van de Par, "Background, concept, and architecture for the recent MPEG Surround standard on multichannel audio compression," *J. Audio Eng. Soc.*, vol. 55, no. 5, pp. 331-351, May, 2007.

[19] M. A. Gerzon, "General metatheory of auditory localization," in *Proc. 92nd Audio Eng. Soc. Conv.*, Vienna, Austria, 1992.

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

[21] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent component analysis*. New York: John Wiley & Sons, 2001.

[22] J. Usher and J. Benesty, "Enhancement of spatial sound quality: a new reverberation-extraction audio upmixer," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 7, pp. 2141-2150, Sep. 2007.

[23] J. Benesty, J. Chen, and Y. Huang, "Binaural noise reduction in the time domain with a stereo setup," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no.8, pp. 2260-2272, Nov. 2011.

[24] J. Thompson, B. Smith, A. Warner, and J. M. Jot, "Direct-diffuse decomposition of multichannel signals using a system of pair-wise correlations," in *Proc. 133rd Audio Eng. Soc. Conv.*, San Francisco, 2012.

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

[26] M. Goodwin, "Geometric signal decompositions for spatial audio enhancement," in *Proc. ICASSP*, Las Vegas, 2008, pp. 409-412.

[27] J. Merimaa, M. M. Goodwin, J. M. Jot, "Correlation-based ambience extraction from stereo recordings", *in 123rd Audio Eng. Soc. Conv.*, New York, Oct. 2007.

[28] 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.

[29] Y. H. Baek, S. W. Jeon, Y. C. Park, and S. Lee, "Efficient primary-ambient decomposition algorithm for audio upmix," in *Proc. 133rd Audio Eng. Soc. Conv.*, San Francisco, 2012.

[30] M. Briand, D. Virette and N. Martin, "Parametric representation of multichannel audio based on principal component analysis," in *Proc. 120th Audio Eng. Soc. Conv.*, Paris, 2006.

[31] J. Se-Woon, H. Dongil, S. Jeongil, P. Young-Cheol, and Y. Dae-Hee, "Enhancement of principal to ambient energy ratio for PCA-based parametric audio coding," in *Proc. ICASSP*, Dallas, 2010, pp. 385-388.

[32] D. Shi, R. Hu, W. Tu, X. Zheng, J. Jiang, and S. Wang, "Enhanced principal component using polar coordinate PCA for stereo audio coding," in *Proc. ICME*, Melbourne, Australia, 2012, pp. 628-633.

[33] J. He, E. L. Tan, and W. S. Gan, "Time-shifted principal component analysis based cue extraction for stereo audio signals," in *Proc. ICASSP*, Vancouver, Canada, 2013, pp. 266-270.

[34] I. Jolliffe, *Principal component analysis, 2nd ed.*. New York: Springer-Verlag, 2002.

[35] C. Uhle, A. Walther, O. Hellmuth, and J. Herre, "Ambience separation from mono recordings using non-negative matrix factorization", in *Proc. 30th Audio Eng. Soc. Int. Conf.*, Saariselka, Finland, 2007.

[36] H. Sawada, S. Araki, R. Mukai, and S. Makino, "Blind extraction of dominant target sources using ICA and time–frequency masking," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 6, pp. 2165–2173, Nov. 2006.

[37] C. Faller and F. Baumgarte, "Binaural cue coding—part II: schemes and applications," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 6, pp. 520–531, Nov. 2003.

[38] J. Blauert, *Spatial hearing: the psychophysics of human sound localization.* Cambridge, MA: MIT Press, 1997.

[39] K. Kinoshita, M. Delcroix, T. Nakatani, and M. Miyoshi, "Suppression of late reverberation effect on speech signal using long-term multiple step linear prediction," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 17, no. 4, pp. 534–545, May. 2009.

[40] E. A. P. Habets, S. Gannot, and I. Cohen, "Late reverberant spectral variance estimation based on a statistical model," *IEEE Signal Process. Letters*, vol. 16, no. 9, pp. 770-773, Sep. 2009.

[41] E. Vincent, R. Gribonval and C. Févotte, "Performance measurement in blind audio source separation" *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 4, pp. 1462-1469, Jul. 2006.

[42] A. Jeffress, "A place theory of sound localization," *J. Comput. Physiol. Psychol.*, vol. 41, no. 1, pp. 35-39, Feb. 1948.

[43] P. X. Joris, P. H. Smith, and T. Yin, "Coincidence detection in the auditory system: 50 years after Jeffress," *Neuron*, vol. 21, no. 6, pp.1235-1238, Dec. 1998.

[44] Y. Ando, and P. Cariani, *Auditory and Visual Sensation*. New York: Springer, 2009

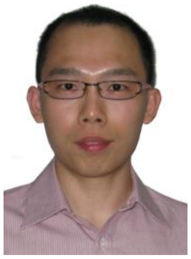
[45] J. Capon, "High resolution frequency wave number spectrum analysis," *Proc. IEEE*, vol. 57, no. 8, pp. 1408–1418, Aug. 1969.

[46] C. Faller, "Parametric multichannel audio coding: synthesis of coherence cues," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 1, pp. 299-310, Jan. 2006.

Jianjun He (S'12) received his BEng degree in Automation from Nanjing University of Posts and Telecommunications, P. R. China in 2011 and is currently pursuing his Ph.D. degree in Electrical and Electronic Engineering in Nanyang Technological University, Singapore. His research interests include: audio and acoustic signal processing, 3D audio, psychoacoustics, and emerging audio applications.

Paper published in IEEE/ACM Trans. Audio, Speech, Lang. Processing, vol. 22. no. 2, pp. 505-517, Feb. 2014

http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6698287



Ee-Leng Tan received his BEng (1st Class Hons) and PhD degrees in Electrical and Electronic Engineering from Nanyang Technological University in 2003 and 2012, respectively. Currently, he is with NTU as a research fellow. His research interests include image/audio processing and real-time digital signal processing.

Woon-Seng Gan (M'93-SM'00) received his BEng (1st Class Hons) and PhD degrees, both in Electrical and Electronic Engineering from the University of Strathclyde, UK in 1989 and 1993 respectively.

He is currently an Associate Professor and the Head of Information Engineering Division, School of Electrical and Electronic Engineering in Nanyang Technological University. His research interests span a wide and related areas of adaptive signal processing, active noise control, and directional sound system. He has published more than 220 international refereed journals and conferences, and has granted five Singapore/US patents. He had co-authored three books on *Digital Signal Processors: Architectures, Implementations, and Applications* (Prentice Hall, 2005), *Embedded Signal Processing with the Micro Signal Architecture*, (Wiley-IEEE, 2007), and *Subband Adaptive Filtering: Theory and Implementation* (John Wiley, 2009).

He is currently a Fellow of the Audio Engineering Society(AES), a Fellow of the Institute of Engineering and Technology(IET), a Senior Member of the IEEE, and a Professional Engineer of Singapore. He is also an Associate Technical Editor of the Journal of Audio Engineering Society (JAES); Associate Editor of the IEEE Transaction on Audio, Speech, and Language Processing (ASLP); and Editorial member of the Asia Pacific Signal and Information Processing Association (APSIPA) Transactions on Signal and Information Processing. He is currently a member of the Board of Governor of APSIPA.