EURASIP Journal on
Audio, Speech, and Music Processing
a SpringerOpen Journal

RESEARCH Open Access

Estimation and quantization of ICC-dependent phase parameters for parametric stereo audio coding

Dong-il Hyun[1*], Young-cheol Park[2] and Dae Hee Youn[1]

Abstract

Conventional parametric stereo (PS) audio coding employs inter-channel phase difference and overall phase difference as phase parameters. In this article, it is shown that those parameters cannot correctly represent the phase relationship between the stereo channels when inter-channel correlation (ICC) is less than one, which is common in practical situations. To solve this problem, we introduce new phase parameters, channel phase differences (CPDs), defined as the phase differences between the mono downmix and the stereo channels. Since CPDs have a descriptive relationship with ICC as well as inter-channel intensity difference, they are more relevant to represent the phase difference between the channels in practical situations. We also propose methods of synthesizing CPDs at the decoder. Through computer simulations and subjective listening tests, it is confirmed that the proposed methods produce significantly lower phase errors than conventional PS, and it can noticeably improve sound quality for stereo inputs with low ICCs.

Keywords: Parametric stereo (PS), Inter-channel phase difference (IPD), Overall phase difference (OPD), Inter-channel correlation (ICC), Channel phase difference (CPD), Stereo audio coding, Spatial audio coding

Introduction

In an effort to efficiently represent multi-channel audio, spatial audio coding (SAC) has been studied extensively during the last decade [1-4]. Among SAC schemes, parametric stereo (PS) [5] drew keen attention due to its simple but effective way of representing stereo audio. PS presents stereo audio as a downmixed mono, together with relevant spatial parameters. Past researches indicate that PS can provide stable stereo quality at bit rates of a few kbps for spatial parameters [5]. After being combined with binaural cue coding (BCC) [6], PS was expanded to multi-channel applications, so that it was adopted in MPEG Surround as a stereo tool [7-10]. PS was also included in HE-AACv2 [9] and the recently developed unified speech and audio coding (USAC) [10] standards.

Parametric representation of stereo sound image can be accomplished by using interaural cues: interaural level difference (ILD), interaural time difference (ITD), and interaural correlation (IC). The direction of sound source can be represented using ILD and ITD, and IC is used to represent the width of the sound source. The PS encoder exploits inter-channel parameters rather than interaural cues because output signals can be transmitted to each ear differently according to the playback system, which can result in different interaural cues. Specifically, for headphone playback, since the transducer output is directly applied to each ear, the inter-channel parameters, such as the inter-channel level difference (ICLD), inter-channel time difference (ICTD), and inter-channel correlation (ICC), can instantly affect the interaural sensations. In PS, the original stereo sound is regenerated from the downmixed mono using these channel parameters. Thus, to obtain the original stereo image with high fidelity, the decoder should properly distribute the channel parameters to the left and right output channels. In the PS decoder, ICLD is always correctly reconstructed because the encoder uses a constraint to limit the gains for each channel. ICTD, however, cannot be correctly reconstructed without *a priori* information of the phase

*Correspondence: fireloaf@dsp.yonsei.ac.kr
[1]Department of Electrical & Electronic Engineering, Yonsei University, Seoul, Korea
Full list of author information is available at the end of the article

distribution over the left and right output channels. In BCC, ICTD is equally distributed over the output channels [6,11]. However, since the channel with the higher energy has a smaller phase difference from the downmixed signal than the other channel, equal distribution of the ICTD parameter can cause degradation of sound quality. As a remedy to this problem, PS adopts the overall phase difference (OPD) as an additional phase parameter.

The practical PS encoder extracts the inter-channel phase difference (IPD) instead of the ICTD, although ICTD is known to be more reliable than IPD in representing spatial characteristics of input audio [5]. ICTD can be analyzed in both the time and frequency domains. In the time domain, the time lag maximizing the cross-correlation between the two channels can be ICTD, but this process demands a considerable amount of computational complexity [6]. ICTD can also be analyzed in the frequency domain by differentiating the phase differences, but this approach often produces inaccurate time delays because of the ambiguity caused by phase wrapping.

Previously, there have been many studies on ICTD and IPD analyses. To solve the phase-wrapping problem of ICTD, the utilization of linear regression was proposed in [11], where the validity of ICTD was also checked by considering ICC. Also, PS employs a frequency domain IPD estimation method that does not require phase unwrapping [5]. In [12], the relationship between OPD and other spatial parameters was mathematically established. It was shown that OPD could be estimated using other spatial parameters, such as inter-channel intensity difference (IID), IPD, and ICC, at the decoder, which resulted in saving bits for OPD quantization. A modified version of OPD estimation proposed in [12] was included in USAC standardization [13].

Errors in IPD and OPD estimation can cause not only distortion of spatial perception, but also deterioration of audio quality [11]. Thus, IPD and OPD analyses should be done with great care. Stereo audio can be separated into primary and ambient components. ICC is relevant only to the highly correlated primary components between channels (such as discrete pairwise-panned instruments), not to the uncorrelated ambient signals (such as reverberation, rain, or applause) [4], and IPD is also associated with the direction of the primary component, which implies that ICC and IPD are mutually dependent and combined in the binaural cues corresponding to the primary components. If that is the case, ICC should be considered for the analysis and synthesis of IPD. Previously, the relation between ICC and ICTD was experimentally analyzed [14]. It was shown that, when ICC was high, ICTD became a relevant cue for the direction of the sound source, and adversely, ICTD was less important when ICC was low.

In this article, we propose improved analysis and synthesis methods for the phase parameters. We first analyze the dependency of IPD on ICC in the process of OPD estimation. Based on the analysis, we propose a new IPD analysis and synthesis method in which IPD is measured dependently on the ICC parameter. Consequently, the proposed method can improve the audio quality, in particular when ICC is low. In this article, the quantization and transmission of the proposed phase parameters are also discussed. Later, we propose methods for estimating the OPD parameters using the other spatial parameters.

The rest of this article is organized as follows. In Section 2, a new phase parameter analysis and synthesis method is proposed and the validity of the parameters is verified in comparison with the conventional methods. Section 3 presents the parameterization of the proposed phase parameters. In Section 4, the overall performance of the proposed phase analysis/synthesis system are measured and compared with the previous methods through objective and subjective tests. Finally, conclusions are drawn in Section 5.

ICC-dependent phase parameters

In this section, the problems with the conventional methods of phase representation are reviewed, and new phase parameters, which can effectively represent the phase information in the stereo input, will be introduced.

Phase parameters in PS

In practical stereo systems, the covariance matrix derived from the two input channels contains most of the salient information. The covariance matrix of the parameter band b can be obtained as

$$\mathbf{W}[b] = \begin{bmatrix} R_{LL}[b] & R_{LR}[b] \\ R_{RL}[b] & R_{RR}[b] \end{bmatrix}, \tag{1}$$

where $R_{IJ}[b] = \sum_{k=k_b}^{k_{b+1}-1} X_I[k] X_J^*[k], I, J = L, R$, k is the frequency bin index and k_b is the start index of the parameter band b. The spatial parameters defined in PS can directly be obtained from the elements of the covariance matrix in Equation (1). IID, ICC, and IPD, respectively, are computed as

$$\text{IID} = 10 \log 10 \left(\frac{R_{LL}}{R_{RR}} \right), \quad \text{ICC} = \frac{|R_{LR}|}{\sqrt{R_{LL} R_{RR}}}, \quad \text{and}$$

$$\text{IPD} = \angle \left(\frac{R_{LR}}{\sqrt{R_{LL} R_{RR}}} \right). \tag{2}$$

It is important to note that ICC and IPD, respectively, are the magnitude and phase of the correlation coefficient between the two input channels, i.e., $\frac{R_{LR}}{\sqrt{R_{LL} R_{RR}}}$

To understand the dependency between IPD and ICC, the cross-channel correlation R_{LR} can be depicted in the complex domain. Consider that R_{LR} is analyzed as

$\sum_{k=k_b}^{k_{b+1}-1} X_L[k]\,X_R^*[k]$. When both channel signals are fully correlated, vectors corresponding to $X_L[k]\,X_R^*[k], k = k_b, k_b+1, \ldots, k_{b+1}-1$, have the same direction, and the overall magnitude is equal to $\sqrt{R_{LL}R_{RR}}$. Thus, ICC becomes 1. Figure 1a shows this case. When both channels are uncorrelated, however, the vectors corresponding to $X_L[k]\,X_R^*[k]$ are in random directions, and thus the overall magnitude is much smaller than 1, which results in small ICCs. Figure 1b shows the second case. A high ICC implies that the primary components are dominant in the channel signals, and thus the IPD is mainly determined by the direction of the primary components. On the other hand, a low ICC implies that the ambient components are dominant and the primary components cannot affect the IPD. Thus, the IPD obtained with low ICC signals does not contain meaningful directional cues, and no phase synthesis at the decoder is desirable. Similar observations can be found in [14], in which the dependency between the IPD and ICC was also stated.

The IPD representing the phase difference between the stereo inputs $X_L[k]$ and $X_R[k]$ can be estimated as [5]:

$$\mathrm{IPD}[b] = \angle\left(\sum_{k=k_b}^{k_{b+1}-1} X_L[k]\,X_R^*[k]\right). \tag{3}$$

The IPD defined in Equation (3) represents the total amount of phase difference between the two input channels. By properly distributing the IPDs over the output channels, the spatial impression of the original stereo signal can be reproduced. A simple approach to the IPD distribution is to divide the total IPD equally in two and apply them to the left and right output channels, respectively. However, this approach cannot guarantee the exact production of the original spatial impression, since the phase difference in this case cannot appropriately represent the spatial attribute of the sound source [6,11]. To solve this problem, the OPD parameter is commonly used for phase synthesis. The OPD representing the phase difference between $X_L[k]$ and the downmixed mono $S[k]$ is formulated as [5]:

$$\mathrm{OPD}[b] = \angle\left(\sum_{k=k_b}^{k_{b+1}-1} X_L[k]\,S^*[k]\right). \tag{4}$$

It is straightforward to show that the OPD and the other primary spatial parameters, such as IID, ICC, and IPD, are related as [12]:

$$\mathrm{OPD}[b] = \angle\left(c[b] + \mathrm{ICC}[b]\,e^{j\mathrm{IPD}[b]}\right), \quad c[b] = 10^{\mathrm{IID}[b]/20}. \tag{5}$$

The relationship in the above equation indicates that an exact OPD can be obtained from IID, ICC, and IPD parameters only if the parameter quantization is not involved. Thus, it can be said that OPD is a redundant parameter. Furthermore, it was shown in [12] that an OPD estimated using the quantized parameters offered similar root mean square (RMS) errors as quantizing the OPD itself, even with fewer bits. The OPD estimation in Equation (5) can geometrically be interpreted in the complex domain, as shown in Figure 2. The circle with the diameter of ICC is a distance of $c[b]$ away from the origin, and the point P is positioned on the circle by the rotation angle IPD. In this diagram, the OPD is considered as an angle between the real axis and the line spanned by the origin and the point P. The dynamic range of the OPD gets narrower as the ICC approaches to zero and $c[b]$ gets larger. On the contrary, when $c[b]$ reaches its minimum (1) and the ICC reaches its maximum (1), respectively, the dynamic range of OPD will increase up to $\pm\pi/2$. Especially, in this extreme case, the OPD varies rapidly when the IPD gets close to π.

In [13], another relationship between the OPD and the other parameters was derived using a geometric represen-

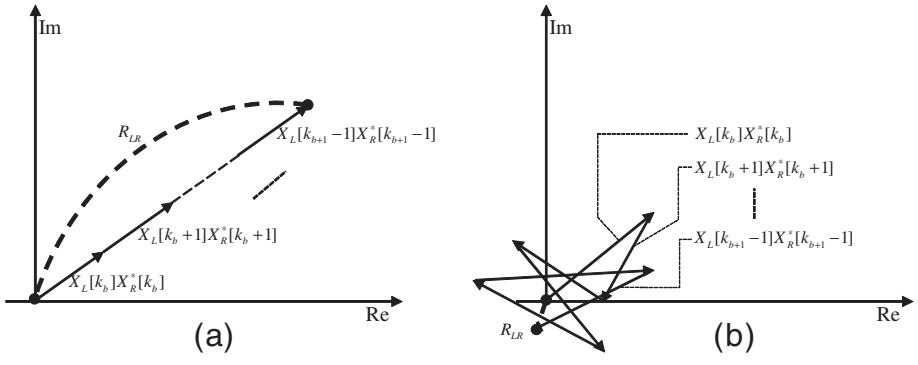

Figure 1 R_{LR} on the complex domain: (a) fully correlated; (b) uncorrelated.

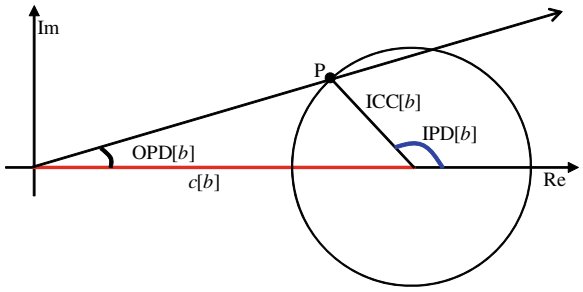

Figure 2 Geometric representation of OPD estimation.

tation of the stereo inputs. According to this approach, the OPD can be expressed as

$$\mathrm{OPD}[\,b] = \arctan\left(\frac{c_2 \sin(\mathrm{IPD}[\,b]\,)}{c_1 + c_2 \cos(\mathrm{IPD}[\,b]\,)}\right),$$
$$c_1 = \sqrt{\frac{10^{\frac{\mathrm{IID}[b]}{10}}}{1+10^{\frac{\mathrm{IID}[b]}{10}}}}, \quad \text{and } c_2 = \sqrt{\frac{1}{1+10^{\frac{\mathrm{IID}[b]}{10}}}}. \tag{6}$$

This method was premised on an assumption that the ICC is 1. Equation (6) can be obtained from Equation (5) using ICC = 1 and $c = \frac{c_1}{c_2}$. Thus, if ICC is 1 for all frames and parameter bands, exact OPDs can be obtained from IID and IPD parameters using Equation (6), which results in bit saving, since we do not need to quantize the OPDs [13]. But if the ICC is not 1, this method may lead to the wrong OPD and, in turn, cause degradation of audio quality, which will be explained in more detail in the next section. The above-mentioned OPD estimation methods were developed using the quantization tables specified in PS [5].

In the conventional PS decoder, the stereo signals are reconstructed from the mono downmix (S) and its decorrelated signal (S_d), using an upmix matrix, as given in [5]:

$$\begin{bmatrix} L' \\ R' \end{bmatrix} = \begin{bmatrix} U_{11} \cdot e^{j\mathrm{OPD}} & U_{12} \cdot e^{j\mathrm{OPD}} \\ U_{21} \cdot e^{j(\mathrm{OPD}-\mathrm{IPD})} & U_{22} \cdot e^{j(\mathrm{OPD}-\mathrm{IPD})} \end{bmatrix} \begin{bmatrix} S \\ S_d \end{bmatrix}, \tag{7}$$

where $U_{11} = c_1 \cdot \cos(\alpha+\beta)$, $U_{12} = c_1 \cdot \sin(\alpha+\beta)$, $U_{21} = c_2 \cdot \cos(-\alpha+\beta)$, $U_{22} = c_2 \cdot \sin(-\alpha+\beta)$, and $\alpha = \frac{1}{2}\arccos(ICC)$, $\beta = \arctan\left(\frac{c_2-c_1}{c_2+c_1}\tan(\alpha)\right)$, respectively. In Equation (7), we omitted the band index b for ease of description. From now on, the band index will not be used except where it is indicated. By separating the OPD from the IPD, Equation (7) can be rewritten as

$$\begin{bmatrix} L' \\ R' \end{bmatrix} = e^{j\mathrm{OPD}} \begin{bmatrix} 1 & 0 \\ 0 & e^{-j\mathrm{IPD}} \end{bmatrix} \begin{bmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{bmatrix} \begin{bmatrix} S \\ S_d \end{bmatrix}. \tag{8}$$

New phase parameters

In practical situations, audio signals are often simply modeled as a sum of the primary and ambient components. In this case, the mono downmix (S) and its decorrelated signal (S_d) in Equation (7) correspond to the primary and ambient components, respectively. In this case, if the ICC is close to 1 it implies that the primary component in the channel signals is dominant, and thus the IPD obtained using the same signals will comprise mainly the directional attribute of the primary component. On the other hand, the IPD in a low ICC situation is easily affected by the strong ambient component, so that it cannot effectively represent the directional attribute of the primary component. Furthermore, directional attributes are often inappropriate in a low ICC situation. However, the upmixing in Equation (8) cannot correctly reflect these observations. The main reason is that the IPD is used to synthesize the phase for the right output channel without consideration of the relationship between the phase and the other spatial parameters, such as the ICC and IID.

To have an exact phase relationship between the left and right channel inputs, we use a method for measuring the two channel phase differences (CPDs), rather than the OPD and IPD. We first define the new CPD parameters as the phase differences between the mono downmix and channel with the higher energy (dominant channel) and the channel with the smaller energy (recessive channel), referred to as CPD1 and CPD2, respectively. Then, these CPD parameters are estimated as

$$\begin{aligned} \mathrm{CPD1}[\,b] &= \angle\left(\sum_{k=k_b}^{k_{b+1}-1} X_L[\,k]\,S^*[\,k]\right) \\ \mathrm{CPD2}[\,b] &= \angle\left(\sum_{k=k_b}^{k_{b+1}-1} S[\,k]\,X_R^*[\,k]\right) \end{aligned}, \quad \text{if IID} \geq 0. \tag{9}$$

IID is positive when the left channel has higher energy than the right channel, and vice versa. Thus, by definition, if IID < 0, CPD1 and CPD2 will interchangeably be defined. Similar to Equation (5), CPD1 and CPD2 can also be expressed using IID, IPD, and ICC parameters

$$\begin{aligned} \mathrm{CPD1} &= \angle\left(c + \mathrm{ICC}e^{j\mathrm{IPD}}\right) \\ \mathrm{CPD2} &= \angle\left(\mathrm{ICC}e^{j\mathrm{IPD}} + \frac{1}{c}\right) \end{aligned}, \quad \text{if IID} \geq 0. \tag{10}$$

Now, using CPD1 and CPD2, the upmix matrix in Equation (8) can be re-written as

$$\begin{bmatrix} L' \\ R' \end{bmatrix} = \begin{bmatrix} e^{j\mathrm{CPD1}} & 0 \\ 0 & e^{-j\mathrm{CPD2}} \end{bmatrix} \begin{bmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{bmatrix} \begin{bmatrix} S \\ S_d \end{bmatrix}, \quad \text{if IID} \geq 0. \tag{11}$$

Also, if IID $<$ 0, then CPD1 and CPD2 in the above equation should be interchanged.

The conventional upmix matrixing in Equation (8) assumes that the sum of the phase difference between the left and right channels is equal to the IPD, and thus the phase difference of the right channel, with respect to the mono downmix, is determined as OPD − IPD. On the other hand, the upmixing in Equation (11) uses independent CPDs. Thus, the total phase difference between the left and right channels is determined as CPD1 + CPD2. Denoting the sum of CPD1 and CPD2 as a phase difference sum (PDS), we have

$$\begin{aligned}\mathrm{PDS} &= \mathrm{CPD1} + \mathrm{CPD2} \\ &= \angle\left(\left(c + \mathrm{ICC}e^{j\mathrm{IPD}}\right)\left(\mathrm{ICC}e^{j\mathrm{IPD}} + \frac{1}{c}\right)\right) \\ &= \angle\left(1 + \left(c + \frac{1}{c}\right)\mathrm{ICC}e^{j\mathrm{IPD}} + \mathrm{ICC}^2 e^{j2\mathrm{IPD}}\right) \\ &= \angle\left(1 + \left(c + \frac{1}{c}\right)\mathrm{ICC}e^{j\mathrm{IPD}} + \mathrm{ICC}^2\left(2\cos^2(\mathrm{IPD})\right.\right. \\ &\quad \left.\left. - 1 + j2\sin(\mathrm{IPD})\cos(\mathrm{IPD})\right)\right) \\ &= \angle\left(\left(\left(c + \frac{1}{c}\right)\mathrm{ICC} + 2\cos(\mathrm{IPD})\mathrm{ICC}^2\right)e^{j\mathrm{IPD}}\right. \\ &\quad \left. + \left(1 - \mathrm{ICC}^2\right)\right).\end{aligned} \tag{12}$$

Similar to Figure 2, Equation (12) can also geometrically be interpreted as Figure 3, where a circle with the radius $(c + 1/c)\,\mathrm{ICC}+2\cos(\mathrm{IPD})\mathrm{ICC}^2$ is a distance $1-\mathrm{ICC}^2$ away from the origin. The IPD can be interpreted as an angle from the center of the circle to the point Q on the circle. The PDS is the angle between the real axis and the line spanned by the origin and the point Q.

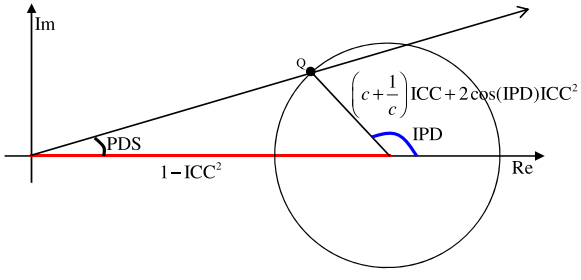

Figure 3 Geometric representation of the relation between the PDS and IPD.

Now, it is straightforward to see that, when ICC = 1, the center of the circle moves to the origin, so that the PDS is equal to the IPD:

$$\mathrm{PDS} = \angle\left(\left(c + \frac{1}{c} + 2\cos(\mathrm{IPD})\right)e^{j\mathrm{IPD}}\right) = \mathrm{IPD}. \tag{13}$$

Also, when ICC = 0, we have PDS $= \angle(1) = 0$.

When the stereo input signals are fully correlated (ICC = 1), the IPD measured using Equation (3) is identical to the total phase difference (PDS). Thus, the assumption premised on the conventional phase synthesis is fully satisfied. When the stereo input signals are uncorrelated (ICC = 0), we have PDS = 0. Thus, no phase needs to be synthesized at the decoder. The IPD, on the other hand, is unpredictable in this case, so that an arbitrary phase difference will be synthesized at the decoder. In addition, as can be seen from Figure 3, $|\mathrm{IPD}| \geq |\mathrm{PDS}|$ for all ICCs, which implies that it is likely to cause excessive phase synthesis only to the right channel because the PS describes the phase of the right channel as OPD − IPD.

Most of these aspects can be resolved using the CPD1 and CPD2 pair defined in Equation (9) instead of the OPD and IPD pair in Equations (3) and (4), as both CPD1 and CPD2 are the relevant parameters dependent on ICC and IID. If the CPD1 and CPD2 pair can exactly represent the phase difference between the left and right inputs, the difference between the PDS and the IPD can be considered as a phase error in the synthesized outputs. For further investigation of this phase error, we plot the PDS versus the IPD according to several IIDs and ICCs, which is shown in Figure 4. The results in Figure 4 were obtained using the quantized IID and ICC values in order to simulate the problem on the decoder side. IIDs of 0, 4, 8, 13, 19, and 30 dB were considered. We also considered non-negative ICCs because PS uses only non-negative ICCs when phase parameters are utilized.

First of all, when the ICC is 1, the IPD is identical to the PDS, regardless of the values of the IID and ICC. When the ICC is close to 1, the IPD roughly matches the PDS in most cases. However, it is noted that when the IID is low (0 dB, for example), even fairly high ICCs produce a significant difference between the PDS and IPD, and the difference becomes insignificant as the IID increases. 0 dB IID corresponds to the case where the sound image is positioned in the median plane, which is very common in practice. Thus, it can be said that, in the conventional PS, a slight decrease of the ICC could result in a significant phase error in the synthesized stereo. It should be noted that, when the IID was 0 dB and the IPD was π (Figure 4a), the PDS always became zero, regardless of the ICC. This is due to the out-of-phase relationship between the channel signals, so that the signals are cancelled out during downmixing. Therefore, a special case should be considered

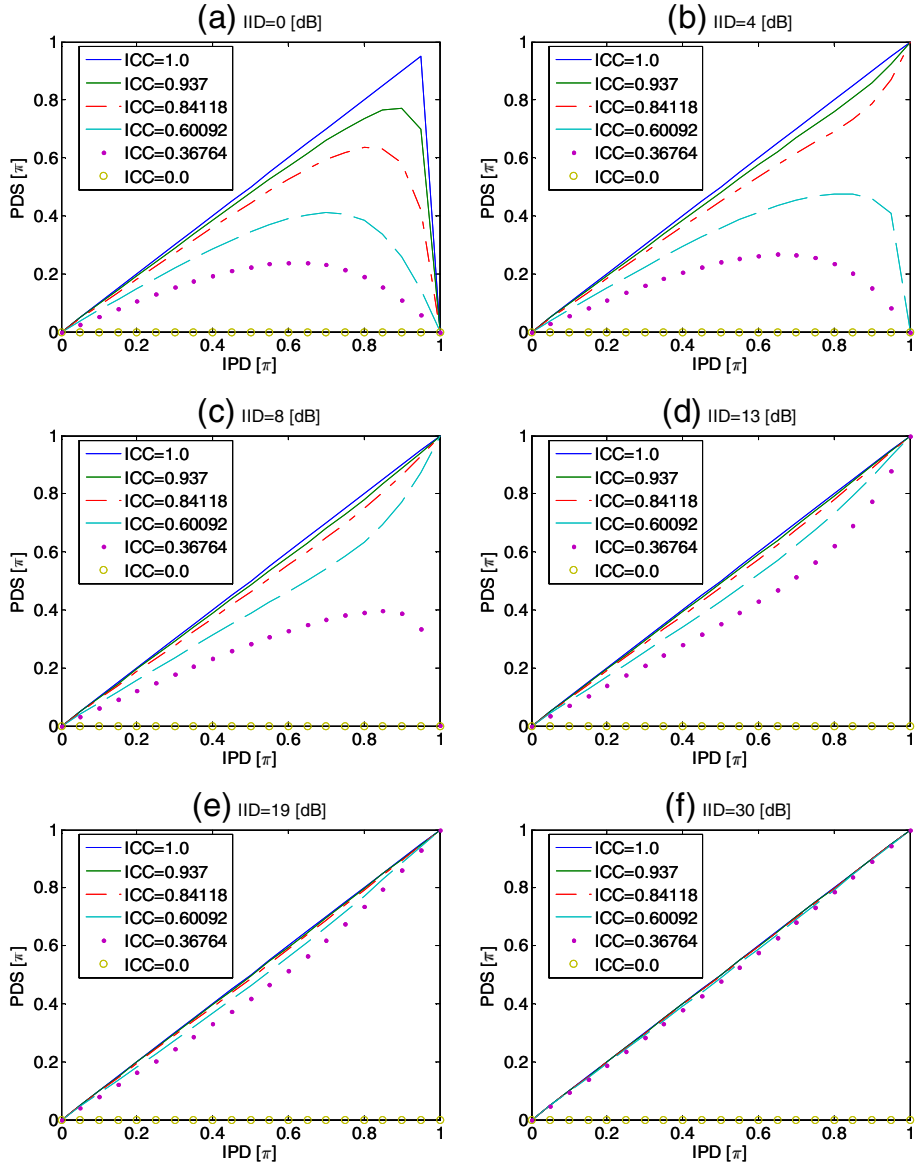

Figure 4 IPD versus PDS according to IID, IPD, and ICC: **(a)** IID = 0, **(b)** IID = 4, **(c)** IID = 8, **(d)** IID = 13, **(e)** IID = 19, and **(f)** IID = 30 dB.

for downmixing when the channel signals have an out-of-phase or near out-of-phase relationship. The downmixing problem is beyond the scope of this article, but the related research has been studied [15,16].

In summary, Figure 4 shows that IPD cannot appropriately represent the phase difference between the left and right channels, and the ICC and IID should be considered when the IPD is used. These results partially agree with the results of the recent research in [14], where it was shown that the relevancy of the ICTD is dependent on the ICC [14]. The ICTD is a valid cue for source localization only when the ICC is larger than a certain threshold. Thus, in [14], the effectiveness of the ICTD was judged by comparing the ICC with a threshold. Analogous to that of [14], the ICC in Figure 4 can be interpreted as a factor for a soft decision.

Based on the observations made for the PDS and IPD, we propose to use the CPD1 and CPD2 pair defined in Equation (9) for the description of the phase difference between the left and right inputs.

Estimation of CPD parameters

For the consideration of the practical PS, where the OPD is not transmitted but estimated at the decoder, we propose two different methods of estimating the parameter pair (CPD1 and CPD2) using IID, IPD, and ICC.

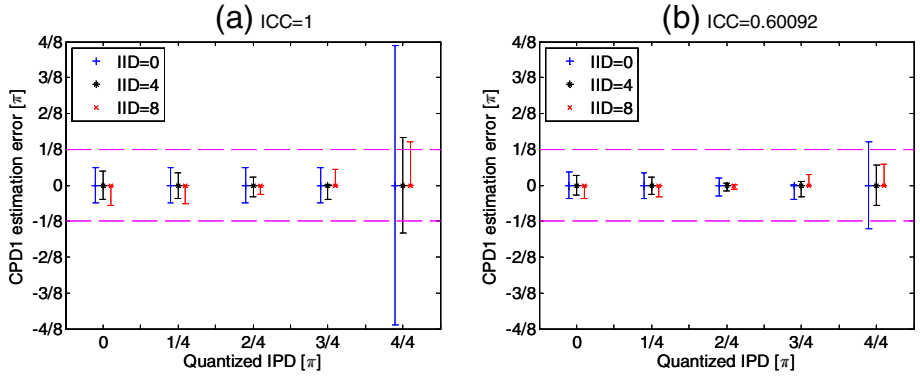

Figure 5 The CPD1 estimation error using the IPD. (a) ICC = 1 and **(b)** ICC = 0.60092.

We redefine the parameter $c[b]$ as $c'[b] = 10^{|IID[b]|/20}$ to discriminate CPDs by the channel energy because the dominant channel is more sensitive to phase error. Now, the CPDs in Equation (10) can be modified as

$$\begin{aligned} \text{CPD1} &= \angle\left(c' + \text{ICC}e^{j\text{IPD}}\right) \\ \text{CPD2} &= \angle\left(\frac{1}{c'} + \text{ICC}e^{j\text{IPD}}\right). \end{aligned} \tag{14}$$

The CPD1 and CPD2 defined in Equation (14) always represent the phase difference for the dominant and recessive channels, respectively. Similar to the conventional estimation method [12], the CPD1 and CPD2 parameters can simultaneously be estimated using IPD, IID, and ICC. Exact CPD values can be recovered when no quantization is involved. However, when the parameters are quantized, the estimated OPDs will contain errors. We measure the errors in the CPD estimation due to the parameter quantization as $\text{CPD1} - \text{CPD1}^{est}$, where CPD1^{est} denotes the estimated CPD1 using the quantized parameters. The estimations errors for CPD1 are displayed in Figure 5. The abscissas of Figure 5a,b is IPDs that were linearly quantized using 3 bits, as in PS. The errors were measured for different IIDs and ICCs. The IID and ICC were assumed to be exactly quantized. Thus, only the IPD quantization was considered. Because IPDs are symmetric about 0, only positive IPD values were used.

The dashed line in the figure indicates the maximum quantization error when CPD1 was directly quantized using 3 bits. If the CPD1 estimation error stays within the dashed line, it can be said that the CPD1 estimation using Equation (14) provides more accurate results than the direct quantization of CPD1. The results in Figure 5 show that, except when the IPD was π, the estimation error is always smaller than the error produced by direct quantization. As the IID increases and the ICC decreases, the variance of the CPD1 estimation error decreases. When IPD = π, there were cases where the CPD1 estimation error was larger than the maximum quantization error. In particular, when ICC = 1 and IID = 0 dB, the CPD1 estimation error was abnormally high. This is the case where the two channel signals were completely out-of-phase. If the estimation error exceeds the quantization error, the OPD estimation in Equation (14) can lead to the degradation of audio quality. To handle the abnormally high estimation error for the out-of-phase signal where the IID = 0 dB,

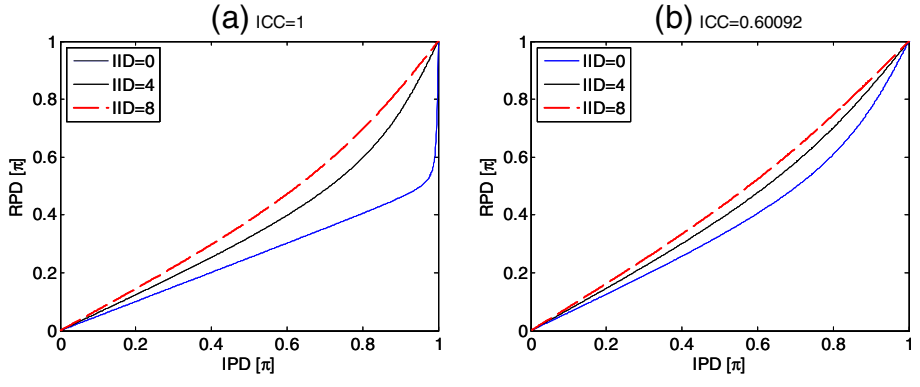

Figure 6 The mapping of IPD to the RPD. (a) ICC = 1 and **(b)** ICC = 0.60092.

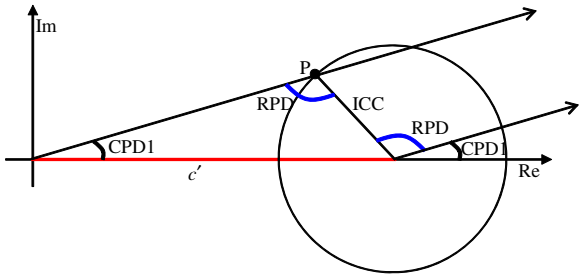

Figure 7 Geometric representation of the CPD1 estimation with the RPD.

ICC = 1, and IPD = π, a nonlinear quantization of the IPD can be used. To implement this, we introduce a new phase parameter, which is referred to as the residual phase difference (RPD) and is defined as

$$\text{RPD} = \text{IPD} - \text{CPD1} \tag{15}$$

The main purpose of introducing the RPD parameter is warping the phase function, so that we can prevent abnormal estimation error especially around π. The relationship between IPD and RPD is plotted in Figure 6. It is shown that Equation (15) nonlinearly maps the IPD on RPD with a higher resolution in the region near π.

Now, a nonlinear quantization of the IPD can be achieved by linearly quantizing the RPD. After quantization, the RPD parameter will be transmitted and the PS decoder will estimate the CPD1 and CPD2 using the RPD, IID, and ICC. To obtain a correct estimation of the CPDs at the decoder, the relationship between the CPDs and the other parameters, including the RPD, should be established. To this end, we can again use the geometrical interpretation in Figure 2. Using the relationship IPD = RPD + CPD1, we can redraw Figure 2 as Figure 7. Then, from Figure 7, we can find the relationship between the OPD and the other parameters

$$\begin{aligned} c'\sin(\text{CPD1}) &= \text{ICC}\sin(\text{RPD}) \\ \text{CPD1} &= \arcsin\left(\frac{\text{ICC}\sin(\text{RPD})}{c'}\right). \end{aligned} \tag{16}$$

The estimation errors of the CPD1 due to quantization of the RPD were measured under the same conditions used in Figure 5, and the results are shown in Figure 8. The abscissas of Figure 8a,b is RPD values that were linearly quantized using 3 bits. The dashed lines again indicate the maximum quantization error that can be obtained when the CPD1 was directly quantized. The results in Figure 8 show that the variance of the estimation error was larger than when estimating the CPD1 using the IPD. However, the error range is still within the maximum quantization error. Furthermore, it is important to note that the CPD1 estimation error for the IPD near π is also within the maximum quantization error.

However, this estimation method has a limitation, in that the CPD2 cannot be estimated at the decoder, since the IPD is not available. This limitation can be overcome using the relationship IPD = RPD + CPD1. Thus, at the decoder, the IPD is re-estimated by summing the transmitted RPD and the estimated CPD1. Finally, the CPD2 is estimated using the obtained IPD.

Performance evaluations

Performance of the proposed phase synthesis method was evaluated by measuring the phase errors and through subjective listening tests. We first measured the errors in the CPD parameters. The proposed CPD estimation methods, based on the IPD and RPD parameters, respectively, were

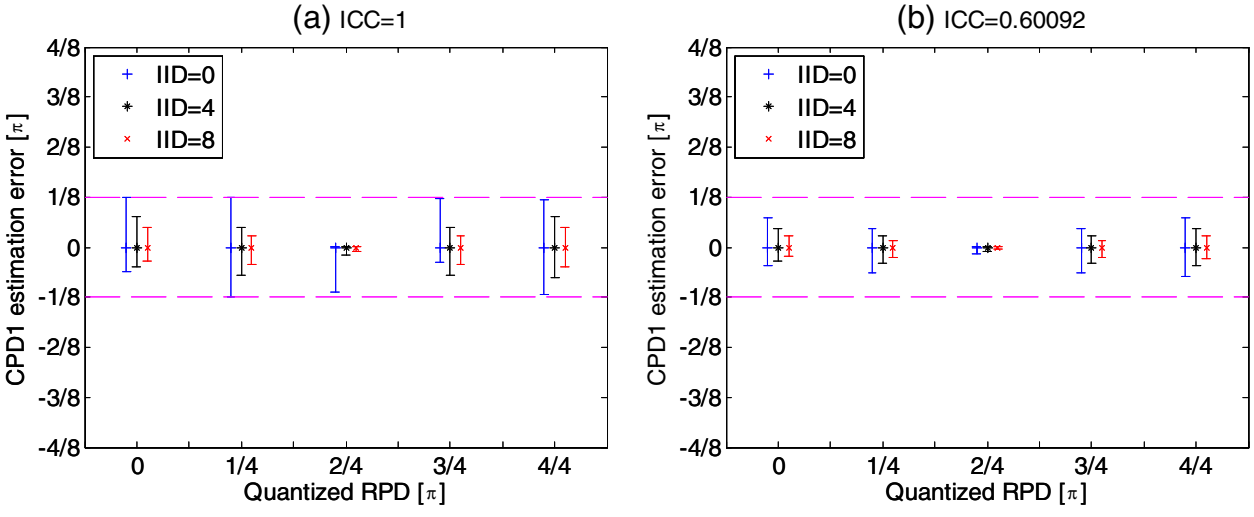

Figure 8 The CPD1 estimation error using the RPD. (a) ICC = 1 and **(b)** ICC = 0.60092.

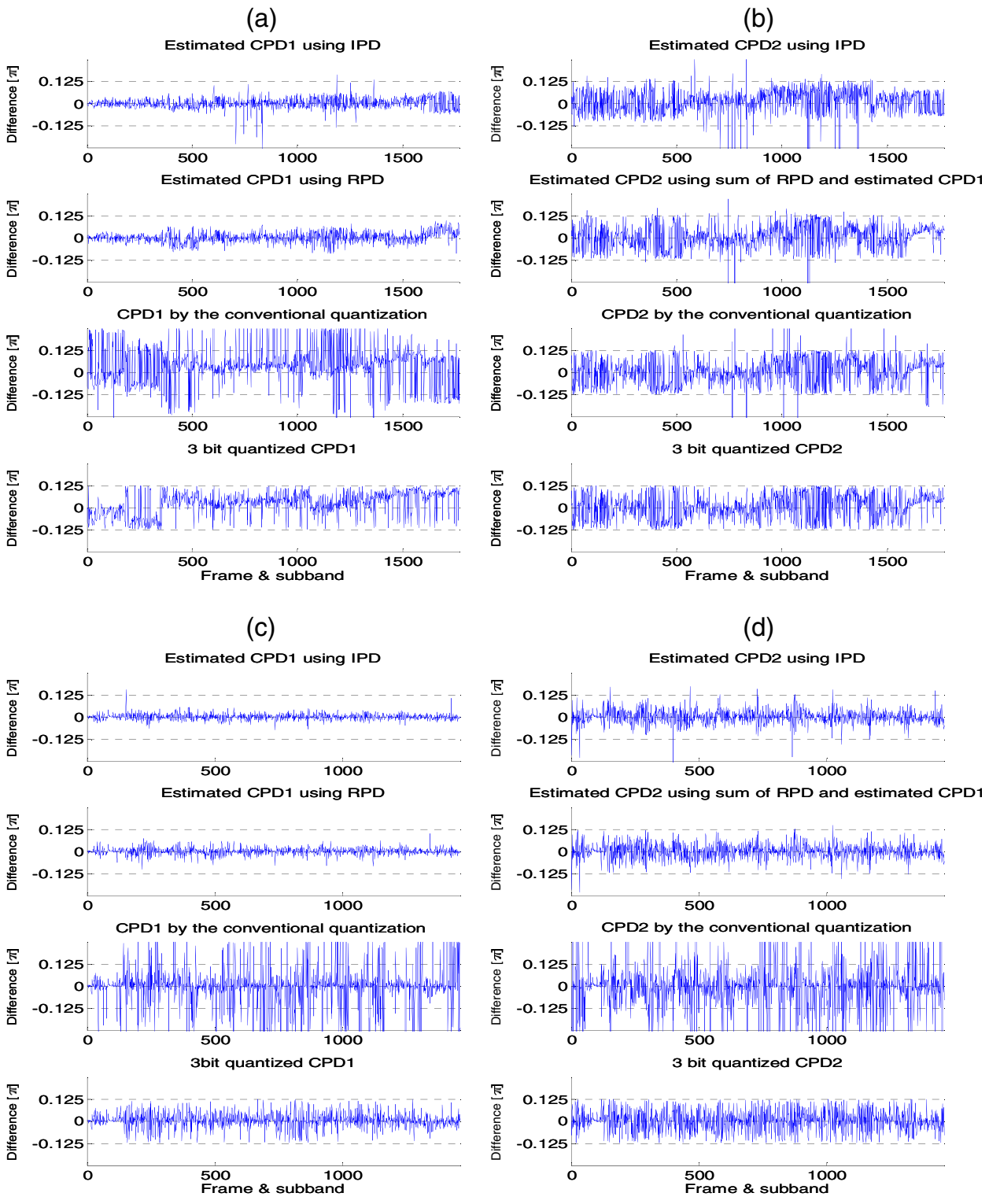

Figure 9 Comparison of phase parameter estimation methods: (a) CPD1 of 'arirang,' (b) CPD2 of 'arirang,' (c) CPD1 of 'speech05,' and (d) CPD2 of 'speech05'.

Table 1 RMS-errors for estimation and quantization methods

	Excerpt					
	Arirang			Speech05		
Method/parameters:	CPD1	CPD2	Total	CPD1	CPD2	Total
IPD-based estimation (Equation 14)	0.0893	0.2260	0.3153	0.0504	0.1222	0.1726
RPD-based estimation (Equation 16)	0.0881	0.2122	0.3004	0.0591	0.1276	0.1867
Conventional quantization (PS)	0.3887	0.2795	0.6682	0.5980	0.6323	1.2304
3-bit quantization (Equation 9)	0.2011	0.2068	0.4078	0.1240	0.1777	0.3016

compared with the method in the conventional PS [5] and with direct quantization. In the direct quantization, it was assumed that the OPD parameters were quantized using additional bits. Thus, the CPD estimation methods are beneficial in terms of bit saving. The hybrid QMF filter adopted in PS [5] was used for the time/frequency representation of the input. The number of hybrid QMF filterbanks and parameter bands were 71 and 20, respectively. The phase parameters were analyzed only for the frequency bands below 2 kHz because it is well known that sound–source localization based on the ITD is dominant at low frequencies. The 2-kHz bandwidth comprises the first 11 parameter bands.

Objective simulation results for estimation methods

Computer simulations were conducted using two stereo excerpts: 'arirang' and 'speech05.' Test excerpt 'arirang' is composed of clean male speech, with channel signals that are near out-of-phase. The other test excerpt, 'speech05,' is composed of male speech with late reverberations that are almost independent. It was assumed that the IPD and RPD were quantized using 3 bits. The IID and ICC were quantized using the quantization tables defined in PS [5]. The measured phase errors are shown in Figure 9. The horizontal axis in the figure represents a merged index of both frame and parameter bands. The dashed line indicates the maximum quantization error of the 3-bit quantizer, which is 1/8.

First, it is shown that the proposed CPD estimation methods (IPD-based and RPD-based) produce much smaller errors in CPD1 than the conventional method used in PS. Furthermore, they are significantly smaller than the maximum quantization error, which shows that the proposed methods are more beneficial because it is possible to obtain a more accurate CPD1 using smaller bits than when quantizing and transmitting the CPD1 itself. In the proposed methods, the CPD1 is associated with the channel with higher energy (dominant channel). Thus, the accuracy of CPD1 is more critical than CPD2 for preserving the original spatial impression. Between the proposed methods, the two methods show a similar degree of estimation errors. However, it should be mentioned that the RPD-based estimation produced smaller peak errors than the IPD-based estimation. With CPD2, the proposed and conventional methods show a similar degree of errors.

The RMS values of the phase errors in Figure 9 are summarized in Table 1. Both the IPD- and RPD-based methods provide significantly smaller RMS errors of CPD1 than direct quantization or the conventional method in PS. For the CPD parameter for the recessive channel (CPD2), the proposed methods show slightly higher RMS errors than direct quantization, and the conventional method shows significantly higher RMS errors than direct quantization. However, since CPD2 is always associated with the channel with lower energy (recessive channel), the errors of CPD2 is perceptually less significant than those of CPD1.

Subjective listening tests

Subjective listening tests were conducted to verify the performance of the proposed parameterization methods. Performances were measured according to the MUSHRA methodology [17]. Hidden anchors were generated by using a low pass filter with a cutoff frequency of 3.5 kHz. The proposed estimation methods, based on the IPD (Equation 14) and RPD (Equation 16), are evaluated, and their performance is compared with the conventional method.

In the proposed methods, the IPD (or RPD), IID, and ICC were quantized and transmitted. The CPD1 and CPD2 were then estimated using the corresponding equations at the decoder. In the conventional PS, the IID, IPD, and ICC were quantized and transmitted, and the OPD were estimated at the decoder. In the direct

Table 2 Test excerpts

Excerpt	Characteristics	Averaged ICC
applaud	Applaud and clapping	0.3505
arirang	Clean male speech	0.9051
motu1	Movie track (the sound of a horse's hoofs)	0.4942
speech60	Male speech with ambience	0.6024
horn30	Brass wind instrument (30-sample ICTD)	1.0000
horn60	Brass wind instrument (60-sample ICTD)	1.0000

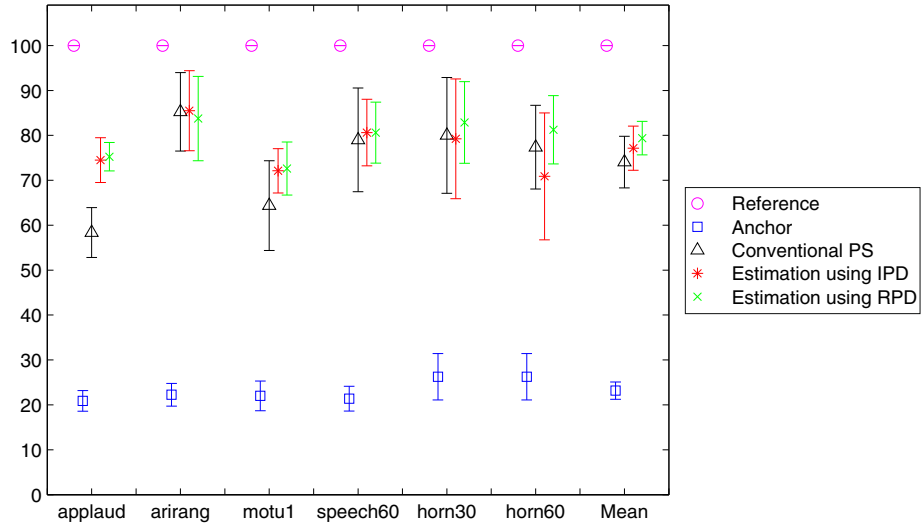

Figure 10 MUSHRA results for phase parameter analysis and synthesis.

quantization method, the IPD, OPD, ICC, and IID were quantized and transmitted. Thus, three more bits were used in comparison with the other methods. To exclude the distortion due to quantization error, the downmixed signal was not quantized.

A subjective listening test was performed with eight subjects experienced in the field of spatial audio. Six test excerpts in Table 2 were presented to the subjects with Sennheiser HD600 headphones. The listening test was conducted using only headphones because inter-channel time or phase differences are irrelevant for loudspeaker playback [6]. In Table 2, the averaged ICCs are also presented. The excerpts were sampled at 44.1 kHz. The test excerpts 'horn30' and 'horn60' were generated by delaying the right channel by 30 and 60 samples, which corresponded to 0.7 and 1.4 ms, respectively. These sample delays can cause phase-reversals for some frequency bands, and as the sample delay gets larger, phase-reversals appear more frequently along the frequency scale.

Since the conventional phase synthesis methods do not consider the other parameters, such as the IID and ICC, quality degradation could be anticipated for test excerpts with low ICCs. The results in Figure 10 are in accordance with the anticipation. The overall qualities of the tested methods were similar to each other for excerpts with relatively high ICCs ('arirang', 'speech60', 'horn30', and 'horn60'). However, for the test excerpt with the lowest ICC, 'applaud,' the proposed methods show significant improvement of sound quality. Thus, it was proven that the proposed methods employing the CPD1 and CPD2 pair can provide better sound quality for stereo inputs with low ICCs than the conventional methods using the IPD and OPD pair.

Between the proposed methods, the RPD-based method scored slightly higher than the IPD-based method for 'horn60.' For 'horn60', the IPD-based method showed a slightly poorer quality than both the conventional and RPD-based methods due to the phase-reversal problem addressed in Section 3. However, the results in Figure 10 show that the problem could be alleviated, and consistent sound quality was obtained using the RPD-based method.

Conclusions

In this article, the problems with conventional phase parameter analysis and synthesis were reviewed, and new phase analysis-synthesis methods, based on new phase parameters, were proposed. It was shown that the assumption for the conventional upmix matrixing was not satisfied in practice because the conventional phase parameters did not consider the relationship between the phase parameter and the other spatial parameters, such as the ICC and IID. It was also shown that a more correct phase representation was possible using the CPD1 and CPD2 pair than using the IPD and OPD pair, and the CPD1 and CPD2 pair could be conveniently synthesized at the decoder. The performance of the proposed methods was evaluated through objective and subjective tests. Test results showed that the proposed methods produced significantly lower phase errors than the conventional methods, and it noticeably improved sound quality for stereo inputs with low ICCs.

Abbreviations

BCC: Binaural cue coding; CPD: Channel phase differences; IC: Interaural correlation; ICC: Inter-channel correlation; ICLD: Inter-channel level difference; ICTD: Inter-channel time difference; ILD: Interaural level difference; IPD: Inter-channel phase difference; ITD: Interaural time difference; OPD: Overall phase

difference; PS: Parametric stereo; RMS: Root-mean-square; RPD: Residual phase difference; SAC: Spatial audio coding; USAC: Unified speech and audio coding.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Department of Electrical & Electronic Engineering, Yonsei University, Seoul, Korea. [2]Computer & Telecommunications Engineering Division, Yonsei University, Wonju, Kangwon, Korea.

Received: 20 May 2012 Accepted: 17 October 2012
Published: 16 November 2012

References

1. KM Short, RA Garcia, ML Daniels, Multi-channel audio processing using a unified domain representation, presented at the 119th Conv. Audio Eng. Soc. New York, NY, USA , October 2005, convention paper 6526
2. M Briand, D Virette, N Martin, Parametric representation of multichannel audio based on principal component analysis, presented at the 120th Conv. Audio Eng. Soc. Paris, France, May 2006, convention paper 6813
3. J Herre, K Kjörling, J Breebaart, C Faller, S Disch, H Purnhagen, J Koppens, J Hilpert, J Röden, W Oomen, K Linzmeier, KS Chon, MPEG surround—the ISO/MPEG standard for efficient and compatible multichannel audio coding. J. Audio Eng. Soc. **56**, 932–955 (2008)
4. M Goodwin, JM Jot, Spatial audio scene coding, presented at the 125th Conv. Audio Eng. Soc. San Francisco, CA, USA, October 2008, convention paper 7507
5. J Breebaart, S van de Par, A Kohlrausch, E Schuijers, Parametric coding of stereo audio. EURASIP J. Appl. Signal Process. **9**, 1305–1322 (2005)
6. F Baumgarte, C Faller, Binaural cue coding—Part I: psychoacoustic fundamentals and design principles. IEEE Trans. Speech Audio Process. **11**(6), 509–519 (2003)
7. ISO/IEC 23003-1:2007, Information technology—MPEG audio technologies—Part 1: MPEG surround, International Standards Organization, Geneva, Switzerland, 2007
8. ISO/IEC 23003-1:2007/Cor.1:2008, Information technology—MPEG audio technologies—Part 1: MPEG surround, TECHNICAL CORRIGENDUM 1, International Standards Organization, Geneva, Switzerland, 2008
9. ISO/IEC 14496-3/Amd.2: 2004, Information technology—Coding of audio-visual objects—Part 3: Audio, Amendment 2: Parametric coding for high-quality audio, International Standards Organization, Geneva, Switzerland, 2004
10. ISO/IEC JTC1/SC29/WG11 N10215, WD on unified speech and audio coding, Busan, Korea, October 2008
11. C Tournery, C Faller, Improved time delay analysis/synthesis for parametric stereo audio coding, presented at the 120th Conv. Audio Eng. Soc. Paris, France, May 2006, convention paper 6753
12. J Lapierre, R Lefebvre, On improving parametric stereo audio coding, presented at the 120th Conv. Audio Eng. Soc. Paris, France, May 2006, convention paper 6804
13. J Kim, E Oh, J Robilliard, Enhanced stereo coding with phase parameters for MPEG unified speech and audio coding, presented at the 127th Conv. Audio Eng. Soc. New York, NY, USA, October 2009, convention paper 7875
14. C Faller, J Merimaa, Source localization in complex listening situations: selection of binaural cues based on interaural coherence. J. Acoust. Soc. Am. **116**(5), 3075–3089 (2004)
15. Samsudin, E Kurniawati, F Sattar, NB Poh, S George, A subband domain downmixing scheme for parametric stereo encoder, presented at the 120th Conv. Audio Eng. Soc. Paris, France, May 2006, convention paper 6815
16. M Kim, E Oh, H Shim, Stereo audio coding improved by phase parameters, presented at the 129th Conv. Audio Eng. Soc. San Francisco, CA, USA, November 2010, convention paper 8289
17. ITU-R BS.1534, Method for the subjective assessment of intermediate quality level of coding systems (MUSHRA), International Standards Organization, Geneva, Switzerland, 2001

doi:10.1186/1687-4722-2012-27
Cite this article as: Hyun *et al.*: **Estimation and quantization of ICC-dependent phase parameters for parametric stereo audio coding.** *EURASIP Journal on Audio, Speech, and Music Processing* 2012 **2012**:27.