Inter-channel transfer function based parametric stereo coding system

Qingbo Huang, Tianshu Qu, Liang Li, Xihong Wu

Key Laboratory on Machine Perception (Ministry of Education), Speech and Hearing Research Center, Peking University, Beijing, China, qutianshu@pku.edu.cn

Abstract

Traditionally, most parametric stereo coding systems use the inter-channel level difference (ILD) and the inter-channel correlation (ICC) as the side information to compress the stereo signal effectively. In this paper, a novel parametric stereo coding method is proposed by using the inter-channel transfer function (ITF) as the side information. The ITF is defined as the transfer function from the mixed signal (the sum of the left channel signal and the right channel signal) to the difference of the left and right channel signal. The ITF contains more spatial information of the stereo sound compared to the ILD and the ICC do. The ITFs of all the frames are grouped together to construct the two dimension matrix. Then, the discrete cosine transform is used to compress the ITFs matrix according to the desired bitrates. Therefore, the redundancies of the stereo signals are diminished not only in the frequency domain but also in the time domain. Lastly, the subjective evaluation experiments based on MUSHRA were carried out to compare the proposed system with the HE-AAC system. The results showed that the proposed system performed comparably to HE-AAC system in the speech signals, the transient musical signals, and the steady state musical signal.

Keywords: Parameter stereo coding, Inter-channel transfer function, HE-AAC, DCT, MUSHRA
Inter-channel transfer function based parametric stereo coding system

1 Introduction

The parametric stereo coding is a technique to efficiently encode stereo audio signals to a mono audio signal with a small amount of spatial parametrical side information which is used for the spatial synthesis process at the decoder side. This coding method enables a high compression ratio by waveform coding on a downmix mono audio signal of the input stereo audio signals [1]. The salient examples include MP3 surround [2], binaural cue coding scheme (BCC) [3,4], the parametric stereo (PS) module [5] in high efficiency-advanced audio coding (HE-AAC) [6,7], and MPEG surround (MPS) [8,9].

In the parametric stereo audio coding method, Inter-channel Level Differences (ILDs) and Inter-channel Correlations (ICCs) are used as the spatial parametric side information. Using ILDs and ICCs as spatial coefficient side information, the coding method has accurate spatial perception as the ILDs and ICCs are directly related to the spatial perception of human auditory. However, the decorrelator, an essential module in the traditional parametric stereo decoder, contains both delays and all-pass filters. The delays are known to have a comb-filter effect which results in an undesirable “metallic” sound, reducing much naturalness of the original signal. The all-pass filters are IIR filters, and have highly nonlinear phase characteristics which lead to clearly noticeable artifacts-like changes in temporal signal structure [10, 11]. Furthermore, the ILDs and ICCs used in the traditional system are only compressed in the frequency domain while the ILDs and ICCs are redundant in the time domain.

In this paper, a novel parameter stereo coding method based on inter-channel transfer function estimation is presented. The outline is as follows. In section 2, the framework of the proposed method is given, and described in detail. In Section 3, the subjective evaluation experiments were carried out based on the multi-stimulus test with hidden reference and anchor (MUSHRA) method. In section 4, the conclusions are drawn based on the evaluation results.

2 The inter-channel transfer function based parametric stereo coding system

The inter-channel transfer function (ITF) is defined as the transfer function from the downmix channel signal to the difference signal of the left and right signals. It contains the relationship between the left channel signal and the right channel signal. Based on the channel transform function estimation, a novel parametric stereo coding scheme will be presented in this section.

2.1 Encoder

The encoder scheme of the proposed parametric stereo coding system is shown in Fig. 1.
Firstly, the left channel signal $L$ and the right channel signal $R$ are fed into the downmix model, where $L$ and $R$ are weighted sum to generate a downmix signal $M$ and a difference signal $D$ as,

$$M = w_l \ast L + w_r \ast R, w_l = w_r = 0.5$$

$$D = w_l \ast L - w_r \ast R, w_l = w_r = 0.5$$

(1)

Secondly, in framing stage, the downmix signal $M$ and the difference signal $D$ are windowed to generate $M_{frame_1}$ to $M_{frame_n}$ frame by frame for signal $M$, and $M_{frame_1}$ to $M_{frame_n}$ frame by frame for signal $D$. $n$ is the total number of frames,

$$n = \left\lceil \frac{m}{l} \right\rceil$$

(2)

where $\left\lceil \cdot \right\rceil$ is the rounding up operator, $l$ is the frame length and $m$ is the length of signal.

Thirdly, each frame of the downmix signal, $M_{frame_i}$, and the corresponding frame of the difference signal $D_{frame_i}$ are fed into the inter-channel transfer function estimation module to estimate the inter-channel transform function $A$. From the definition of inter-channel function, it is obtained that,

$$D_{frame_i} = D_{frame_i} \ast A_i$$

(3)

where $\ast$ is the convolution operator. Expand equation (3) to the matrix form,
\[
\begin{pmatrix}
    d_{k,1} \\
    d_{k,2} \\
    \vdots \\
    d_{k,n}
\end{pmatrix} =
\begin{pmatrix}
    m_{k,1} & m_{k-1,1} & \cdots & m_{k-1,1-p+2} & a_{k,1} \\
    m_{k,2} & m_{k,1} & \cdots & m_{k-1,1-p+3} & a_{k,2} \\
    \vdots & \vdots & \ddots & \vdots & \vdots \\
    m_{k,n} & m_{k,n-1} & \cdots & m_{k-1,1-p+1} & a_{k,n}
\end{pmatrix}
\] (4)

Because \( n \) is much greater than \( p \), Equation (4) is an over determined equation. The least square method is used to get the solution.

In the ITF transfer function matrix compression and quantization model, the matrix \( ITFM \) is constructed by Vector \( A_k \) at all the frames, \( k = 1 \ldots n \),

\[
ITFM = (A_1, A_2, \ldots, A_n)
\] (5)

Then the two-dimension discrete cosine transform (DCT) is used to compress the \( ITFM \) matrix in time domain as well as in frequency domain.

\[
ITFMDCT = DCT2(ITFM)
\] (6)

The \( ITFMDCT \) coefficients with small value have been thrown out and the left \( ITFMDCT \) coefficients are quantized to \( QITFMDCT \) according to the objective bitrate.

Lastly, the quantized DCT coefficient \( QITFMDCT \) is fed to the entropy coding stage to get the side information which will generate bitstream along with the mono signal finally.

### 2.2 Decoder

The decoder scheme of the proposed parametric stereo coding system is shown in Fig. 2.

Fig. 2 The decoding scheme of the proposed parametric coding system.
Firstly, the bitstream is demultiplexed into the side information and the downmix audio signal $M$ in the bitstream demultiplex stage.

Secondly, in the framing stage, the downmix signal $M$ is windowed into frames with $l$-dimension vectors from $M_{\text{frame}_1}$ to $M_{\text{frame}_n}$ frame by frame.

Thirdly, the side information is fed to the entropy decoding stage to get the $\text{QITFMDCT}$. Since the entropy decoding is lossless, $\text{QITFMDCT}$ is the same as it in encoder process

In the Inter-Channel transform function decompression and dequantization stage, the $\text{QITFMDCT}$ is dequantized to get $\text{ITFM'}$. Then the matrix $\text{ITFM'}$ will be got by inverse discrete cosine transform to the reconstructed $\text{ITFM'}$

$$\text{ITFM'} = \text{IDCT}2(\text{ITFM'}_{\text{DCT}})$$  \hspace{1cm} (7)

From the signal $\text{ITFM'}_{\text{DCT}}$, the channel transform functions of each frame, $a'_1, a'_2, \ldots, a'_n$, are got as,

$$\text{ITFM'} = (A'_1, A'_2, \ldots A'_n)$$ \hspace{1cm} (8)

where $n$ is the total number of frames

In upmix stage, the right channel signal $R$ in the $k$th frame is reconstructed by multiplying $M$ and $A$.

$$\begin{pmatrix} d_{k,1} \\ d_{k,2} \\ \vdots \\ d_{k,l} \end{pmatrix} = \begin{pmatrix} m_{k,1} & m_{k-1,1} & \cdots & m_{k-1,j-p+2} \\ m_{k,2} & m_{k,1} & \cdots & m_{k-1,j-p+3} \\ \vdots & \vdots & \ddots & \vdots \\ m_{k,l} & m_{k,l-1} & \cdots & m_{k,j-p+1} \end{pmatrix} \begin{pmatrix} a'_{k,1} \\ a'_{k,2} \\ \vdots \\ a'_{k,p} \end{pmatrix}$$ \hspace{1cm} (9)

where $a'_k = (a'_{k,1}, a'_{k,2}, \ldots, a'_{k,p})^T$ is the channel transform function in the $k$th frame. $d'_k = (d'_{k,1}, d'_{k,2}, \ldots, d'_{k,p})^T$ is the difference channel signal in the $k$th frame and $m_{x,y}$ is the $y$th point of the downmix signal $M$ in the $x$th frame.

Lastly, the left channel signal $L$ and right channel signal $R$ can be obtained by
\[ L_k = M_k + D_k \]  \hspace{1cm} (10)

\[ R_k = M_k - D_k \]  \hspace{1cm} (11)

where \( k \) is the index of the frame number.

### 3 Subjective evaluation experiments

The multi stimulus test with hidden reference and anchor (MUSHRA) is recommended to evaluate the intermediate audio quality and gives accurate and reliable results [12]. To verify the effects of the performance of the proposed system with the parametric stereo module (PS) in HE-AAC [5], a series of subjective listening tests were carried out based on the MUSHRA method.

The AAC module in Nero HE-AAC system [13] is used as the mono-audio coder and decoder either in the proposed system or in the MPEG system. The MUSHRA tests were performed by 12 experienced listeners, age from 24 to 27 years old, 8 males and 4 females. Nine acoustic stereo signals from the MPEG-4 HE-AAC stereo verification test [14] were used as the testing signals. These signals are divided into 3 categories, speech, transient music and steady state music. There are 3 different signals in each category. All the signal content types are described in Table 1. The sample rate of these test signals is 44.1kHz.

<table>
<thead>
<tr>
<th>Type</th>
<th>Content</th>
<th>Duration[s]</th>
</tr>
</thead>
<tbody>
<tr>
<td>speech signals</td>
<td>Suzanne Vega</td>
<td>10.734</td>
</tr>
<tr>
<td></td>
<td>Male German Speech</td>
<td>8.599</td>
</tr>
<tr>
<td></td>
<td>Female English Speech</td>
<td>7.604</td>
</tr>
<tr>
<td>transient music</td>
<td>Harpsichord</td>
<td>7.995</td>
</tr>
<tr>
<td></td>
<td>Castanets</td>
<td>7.725</td>
</tr>
<tr>
<td></td>
<td>Plucked Strings</td>
<td>13.985</td>
</tr>
<tr>
<td>steady state music</td>
<td>Trumpet</td>
<td>10.968</td>
</tr>
<tr>
<td></td>
<td>Pitch Pipe</td>
<td>27.887</td>
</tr>
<tr>
<td></td>
<td>Bagpipe</td>
<td>11.148</td>
</tr>
</tbody>
</table>

A pair of headphones (HD650, Sennheiser) were used to present the testing signals. Four systems were chosen in the evaluation experiments. They are the proposed system, the HE-AAC system, the Hidden reference (original item), and the bandwidth-limited anchors (3.5kHz low-passed item). The experiments were carried out in a quiet room. The bitrate is set at 64kbps. The spatial coefficient side information is 3.45kbps and the compressed mono signal is 60.55kbps.
In this experiment, the MUSHRA scores for average and 95% confidence intervals are presented in Fig. 3-5.

**Fig. 3** The MUSHRA score for the test speech signals

**Fig. 4** The MUSHRA score for the test transient music test signals
Experiment data was analyzed by SPSS to show whether the difference is significant. Fig. 3 shows the MUSHRA score for three speech test signals and the average score. The results show that the proposed system performs better than HE-AAC system at average, but not significantly (p = 0.010). Fig. 4 shows the MUSHRA score for three transient music test signals and the average score. The results show that the HE-AAC system performs better than the proposed system at average, but not significantly (p = 0.234). Fig. 5 shows the MUSHRA score for three steady state music test signals and the average score. The results show the proposed system performs better but not significantly than the HE-AAC system (p = 0.048).

4 Conclusion

The Parametric stereo coding method is an important technology for compressing multichannel acoustic signals. Most parametric stereo coding systems use IIDs and ICCs as the spatial parametric side information to reconstruct the original stereo sound. This paper presents a novel parametric stereo coding method, in which the inter-channel transfer function (ITF) is used as the spatial parametrical side information. In this method, the ITF is estimated using LS method firstly, then the ITFs of all the frames are combined into a two-dimension matrix. Lastly, the two-dimension matrix is compressed by DCT transform. The ITFs preserve the spatial perception cues of the original signal as well as the IID and the ICC cues. The DCT is used to compress the ITFs not only in the frequency domain, but also in the time domain. The subjective evaluation experiments (MUSHRA) are carried out to compare the proposed method with the HE-AAC system. The results showed that the proposed system performed comparably to HE-AAC system in the speech signals, the transient musical signals, and the steady state musical signal.
Acknowledgments

This work was supported by the National High Technology Research and Development Program of China (863 Program: 2015AA016306), and the National Natural Science Foundation of China No.61175043, No.61421062, No. 31170985).

References


