Artificial Stereo Extension of Speech Based on Inter-Channel Coherence

Nam In Park and Hong Kook Kim
School of Information and Communications
Gwangju Institute of Science and Technology (GIST), Gwangju 500-712, Korea
{naminpark, hongkook}@gist.ac.kr

Abstract. In this paper, an artificial stereo extension method of speech is proposed to control the width of auditory image. The proposed stereo extension method employs a decorrelator to make stereo-channel signals without any additional information from mono speech signal. In other words, uncorrelated signal is first generated from mono speech signal passing through the decorrelator, and it is added to or subtracted from the mono signal to obtain the left or right channel signal, respectively. Here, the degree of inter-channel coherence is controlled to vary the width of auditory image. The performance of the proposed stereo extension method is evaluated using a MUSHRA test and compared with that of a parametric stereo method. It is shown from the test that the mean opinion score of the signals extended by the proposed stereo extension method is 92, which is around 5% higher than that of the parametric stereo method.

Keywords: Artificial stereo extension, inter-channel coherence (ICC), decorrelator, auditory image, parametric stereo

1 Introduction

A wide range of multimedia technologies are rapidly being developed for use on portable digital devices such as mobile phones and videophones. With these recent developments, there has also been an increased demand for stereo audio processing on smartphones [1]. Since most speech processing techniques deal only with mono signals, stereo extension of speech can be an alternative to improve the auditory perception of speech.

In this paper, we propose an artificial stereo extension method of speech, which does not require any additional information, allowing the proposed method to be applied as a post-processing step for any other speech processing technique. To this end, the proposed stereo extension method employs a decorrelator to obtain the stereo signals. In other words, an uncorrelated signal is derived from a given mono signal by the decorrelator, and it is added to or subtracted from the mono signal to obtain the left or right channel signal, respectively. Here, the degree of inter-channel coherence (ICC) can be controlled to vary the perceived width of the auditory image. The performance of the proposed stereo extension method is evaluated using a subjective test and compared with that of a conventional parametric stereo method.
The remainder of this paper is organized as follows. Following this introduction, Sections 2 and 3 describe parametric stereo coding and the proposed stereo extension method, respectively. Section 4 discusses the performance of the proposed stereo extension method. Finally, Section 5 concludes this paper.

2 Parametric Stereo Method

Parametric stereo (PS) has recently proven to be a vital component in stereo speech, since it considerably enhances stereo sound performance by encoding the spatial information as additional information [2]. In a complete PS method, the underlying mono signal is carried together with a small amount of additional spatial information extracted by the PS method [2]. Fig. 1 shows a procedure of a PS method. The PS method downmixes stereo input signals into a mono signal and extracts spatial parameters through spatial analysis. The stereo output signals are regenerated by spatial synthesis with the mono signal and spatial parameters.

Fig. 1. Procedure of a parametric stereo method [2].

3 Proposed Stereo Extension Method

Fig. 2 shows a signal flow graph of the proposed stereo extension method to generate artificial stereo signals. As shown in the figure, extended stereo-channel signals, \( x_L(n) \) and \( x_R(n) \), can be obtained as [5]

\[
x_i(n) = g_m(ICC) \cdot x_m(n) + g_d(ICC) \cdot (x_m(n) \ast d(n))
\]  

(1)
\[ x_r(n) = g_m(ICC) \cdot x_m(n) - g_d(ICC) \cdot (x_m(n) \ast d(n)) \]  
(2)

where \( x_m(n) \) is a mono signal and \( d(n) \) is the impulse response of a decorrelator. In addition, \( \ast \) means a convolution operation.

Next, in order to control the degree of the inter-channel coherence (ICC) between the stereo-channel signals, the mono signal, \( x_m(n) \), and the decorrelated signal, \( x_m(n) \ast d(n) \), are weighted by scale factors. In other words, as described in Eqs. (1) and (2), \( g_m(ICC) \) and \( g_d(ICC) \) are multiplied to \( x_m(n) \) and \( x_m(n) \ast d(n) \), respectively. Here, it is required that the two scale factors should be correlated with the ICC. We experimentally show the relationship between ICC and the scale factors, as shown in Table 1. In the table, the ICC is defined as

\[
ICC = \frac{\sum_{n=0}^{N-1} (x_i(n) \cdot x_r(n))}{\sqrt{\sum_{n=0}^{N-1} x_i^2(n) \cdot \sum_{n=0}^{N-1} x_r^2(n)}}
\]  
(3)

where \( N \) is the number of samples used for the ICC computation. As a result, the stereo-channel signals obtained by using Eqs. (1) and (2) has different width of auditory image according to the combination in Table 1.

Table 1. Relationship between ICC and scale factors to provide different width of auditory image in the stereo signals.

<table>
<thead>
<tr>
<th>ICC</th>
<th>0.75</th>
<th>0.5</th>
<th>0.25</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>( g_m(ICC) )</td>
<td>0.94</td>
<td>0.87</td>
<td>0.79</td>
<td>0.71</td>
</tr>
<tr>
<td>( g_d(ICC) )</td>
<td>0.35</td>
<td>0.5</td>
<td>0.61</td>
<td>0.71</td>
</tr>
</tbody>
</table>

4 Performance Evaluation

In order to demonstrate the effectiveness of the proposed artificial stereo extension method, a multiple stimuli with hidden reference and anchor (MUSHRA) listening test [4] was carried out. For this test, six speech files were taken from the sound quality assessment material (SQAM) [5], where they were recorded at a sampling rate of 44.1 kHz with stereo, thus they were down-mixed into mono signals and down-sampled to 16 kHz with mono. Next, two anchors with cut-off frequencies of 3.5 and 7 kHz were prepared. Seven people with no auditory diseases participated in this MUSHRA test. Each participant was presented with the six stimuli and asked to provide a score between 0 and 100, depending upon their opinion of the stereo quality. Here, the ICC of the PS and proposed stereo extension method were set to 0.

Fig. 3 compares the MUSHRA scores, where each column corresponds to the opinion score averaged over seven listeners and all six speech files. The vertical line on the top of each bar denotes the standard deviation of the opinion score. It was shown...
from the figure that the proposed method achieved an average score of 92, which was higher than that of the PS method.

Fig. 3. Comparison of MUSHRA scores

5 Conclusion

In this paper, we proposed an artificial stereo extension method based on ICC for generating stereo signals with improved stereophonic effects. The proposed stereo extension method incorporated a decorrelator to generate uncorrelated signals. After that, stereo signals were generated by controlling on the degree of ICC. It was shown from the subjective listening test that the proposed stereo extension method relatively improved the subjective quality of 5%, compared to a parametric stereo method that required additional information for generating stereo signals.

6 Acknowledgements

This work was supported in part by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2012-010636).

References

2. 3GPP Ts 26.401.: Enhanced accPlus general audio codec; General Description (2008).