A Frequency Domain Nonlinearity for Stereo Echo Cancellation

Ming Wu†, Zhibin Lin†, and Xiaojun Qiu†, Nonmembers

SUMMARY This letter proposes a novel nonlinear distortion for the unique identification of receiving room impulses in stereo acoustic echo cancellation when applying the frequency-domain adaptive filtering technique. This nonlinear distortion is effective in reducing the coherence between the two incoming audio channels and its influence on audio quality is inaudible.

key words: acoustic echo cancellation, nonlinearity, psychoacoustic, stereo

1. Introduction

The stereo acoustic echo cancellers (AECs) are much more complicated than the monophonic ones due to the linear relation between two channels, which leads to the well-known non-uniqueness problem. One successful solution to the non-uniqueness problem is to deliberately add a small amount of nonlinear distortion to each channel e.g. through half-wave rectification [1]. In order to maintain transparency for speech signals, the nonlinear distortion with the adopted function and the parameter \( \alpha \) controls the amount of distortion added. Here \( N \) is assumed to be even for convenience.

where the symbol “c” denotes conjugation, \( f \) is an arbitrary function and the parameter \( \alpha \) controls the amount of distortion added. Here \( N \) is assumed to be even for convenience.

2. Problems and Solutions

2.1 System Description

The stereo echo cancellation system is shown in Fig. 1, where \( g_1 \) and \( g_2 \) are the impulse responses in the transmission room and \( h_1(n) \) and \( h_2(n) \) are the impulse responses in the receiving room. \( x'_1 \) and \( x'_2 \) are the two distorted signals, being obtained by applying the nonlinear transformation to the two incoming transmission room signals \( x_1 \) and \( x_2 \). \( y(n) \) is the echo signal, \( e(n) \) is the error signal and \( y'(n) \) is the output of the adaptive filter. \( w_1(n) \) and \( w_2(n) \) denote the adaptive FIR filters.

2.2 New Nonlinear Transformation

For channel \( i \), the transformed signal \( X'_i(n) \) is obtained as following:

\[ X'_i(k) = X_i(k)e^{j\alpha f(k)}, \quad k = 1, \ldots, N/2 - 1 \]
\[ X'_i(k) = X'_i(N-k), \quad k = N/2 + 1, \ldots, N - 1 \]

where \( X_i(k) = X_i(n) \) is the frequency vector of the original signal \( x_i(n) \).

From above, it can be seen that if \( \alpha \) not very big, the magnitude spectrum of \( x'_i(n) \) is exactly the same with the original signal \( x_i(n) \), so the transformation has little effect on stereo perception. In this paper, for frame \( p \) the adopted function \( f(k) \) is
where $v$ is a zero-mean random noise ranged $[-\pi, \pi]$, $M$ is the number of frequency vectors to be distorted in one block and $j = p \mod (N/2M)$ decides which frequency vectors would be distorted. In addition, only several frequency vectors of original signal are distorted in one block, so when the frequency-domain adaptive algorithm is used, only the coefficients of the adaptive filter corresponding to the distorted frequency vectors need to be updated. When $M = N/2$, all of the original frequency vectors are distorted in one block, which is named fullband distortion. Since the frequency-domain adaptive algorithm is adopted, no additional computation is needed in step 1.

3. Simulations

The validity of the proposed nonlinear distortion was confirmed through the following computer simulation. The source in the transmission room was a 10-s speech stored as 16-bit PCM with a sampling rate of 16 kHz. The four impulse responses $g_1$, $g_2$, $h_1$ and $h_2$ of length 4096 were measured in the listening room of Nanjing University. The frequency-domain adaptive algorithm proposed by Benesty [5] was used to update the weights of $w_1$ and $w_2$ with the adaptive step size $\mu = 0.3$. The length of block $N$ is 1024 and the adaptive filter $w_1$ and $w_2$ are all 1024 taps. The proposed fullband distortion and half-wave rectification distortion ($\alpha = 0.3$) are compared below. Figure 2 shows that after system converges, the misalignment is nearly $-25$ dB with the half-wave nonlinear distortion while with the proposed nonlinear process it is $-25$ dB for $\alpha = 0.03$, $-31$ dB for $\alpha = 0.07$.

Fourteen subjects took part in the psychoacoustic listening experiment to test the effect of the distortion on perception. A MOS (ITU-T recommendation P.800) test was used to identify the listening-quality scale. Experimental results for speech and for music are summarized in Fig. 3. The average response rating is shown on the horizontal axis, and the two types of source material are shown on the vertical axis. The horizontal bars indicate 95% confidence intervals. For speech, the three distortions all appear to be satisfactory. But for music, the proposed distortion is much better than half-wave rectification distortion. Also, preliminary psychoacoustic experiments have shown that the stereophonic spatial localization is not affected with the proposed distortion.

4. Conclusions

In this letter, the distortion integrated with frequency-domain adaptive algorithm is proposed to reduce the coherence of two coming signals. This distortion is hardly detected by human ear due to its less sensibility to phase. Also the misalignment is also less than using the half-wave rectification distortion.

Acknowledgments

Project 10304008 supported by NSFC. The work was also sponsored by SRF for ROCS, SEM and SRFDP.

References
