Karan Pareek, Chengze Zou, Aneesh Athrey
3D Audio
MPATE-GE 2613
Dr. Agnieszka Roginska
28th November, 2018

An Implementation of Crosstalk Cancellation Methods

1 Introduction

Crosstalk cancellation allows for the reproduction of binaural audio through loudspeakers by using digital filters. This project aims to cover and analyze two widely implemented crosstalk cancellation systems, one that uses ridge regression and a regularization factor, and another that involves recursive cancellation using time difference analysis and head shadow coefficients. Since these methods make use of different approaches for achieving crosstalk cancellation, a section has been added to the study that conducts a spectral comparison of the two approaches.

2 XTC using Inverse-Filtering

Efficiency in crosstalk cancellation systems involves delivering a binaural recording over loudspeakers such that the sound intended for the left ear is received by the left ear only, and vice versa. In his book on digital audio processing, Zolzer (2002) implements and discusses a crosstalk cancellation system over loudspeakers. Given a binaural signal $x(n)$, the primary goal of the system is to deliver high fidelity audio that cancels out any form of crosstalk such that a signal is received at the ear canals. Since cross-talk naturally exists in loudspeaker reproduction, the input binaural audio needs to be processed using digital filters. The goal of these filters is to ensure that the sound intended for the left or right ear is delivered to that particular ear while avoiding any form of crosstalk. Given these factors, a simplified form of the crosstalk cancellation system is as follows:

$$d(z) = C(z)H(z)x(z) \tag{1}$$

Here, $C(z)$ is the impulse response of the speakers at the entrance of the ear canals while $H(z)$ contains elements that filter out the audio and perform the required crosstalk cancellation operation. The two elements can be expressed using matrices:

$$C(z) = \begin{pmatrix} C_{11}(z) & C_{12}(z) \\ C_{21}(z) & C_{22}(z) \end{pmatrix} \tag{2}$$

$$H(z) = \begin{pmatrix} H_{11}(z) & H_{12}(z) \\ H_{21}(z) & H_{22}(z) \end{pmatrix} \tag{3}$$

An ideal crosstalk cancellation implementation implies that $x(z) = d(z)$. This is true only if $H(z) = C(z)^{-1}$. However, such conditions are often impossible to achieve. In such a situation, a form of regularization is calculated using ridge regression methods, where the optimal value of $H(z)$ is obtained. Given the inverse condition, the value of the cancellation filter is as follows:

$$H_{opt}(z) = [C^T(z^{-1}C(z) + \beta I)]^{-1}C^T(z^{-1})z^{-m} \tag{4}$$

where β can be defined as a regularization factor and z^{-m} denotes a delay of m samples. One of the challenges in implementing this crosstalk system revolves around determining an ideal value of β. Selecting a low value of β will cause sharp peaks in the spectrum of the processed output while a high value of β will be result in a less accurate filter.

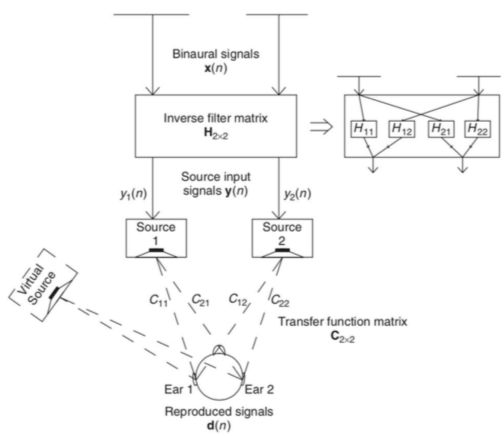

Figure 1: XTC using Inverse-Filtering (Zolzer 2002)

An implementation of the model described above can be achieved in a sequence of steps. First, the impulse responses of the speakers, denoted by C , are transformed into the frequency domain using the Fourier Transform. Given the value of C, the optimal value of H can be derived simply by using the ridge regression method defined in the section above. Once the value of the crosstalk filter is derived, the results are transformed back into the time domain using the Inverse Fourier Transform, before a “a circular shift of half of the applied time-window length is implemented on the inverse filters.” (Zolzer, 2002)

3 XTC using Recursive Filtering

Another method of crosstalk cancellation involves the use of recursion cancellation between the ipsilateral and the contralateral ear. The following diagram is a pivotal step in determining the head shadow coefficients:

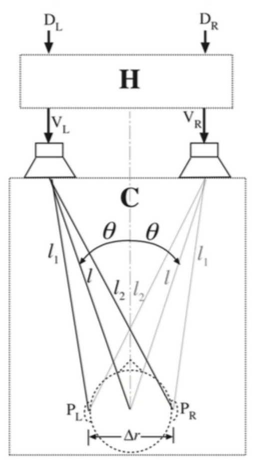

Figure 2: XTC using Recursive Filtering (Roginska et al. 2018)

Where,

r : radius of the head

l : distance to the center of the head from the speaker

l_1 : distance to the ipsilateral ear from the speaker

l_2 : distance to the contralateral ear from the speaker

θ : angle between angle vector connecting each speaker to the center of the head

This method of crosstalk cancellation builds up on the modelling in Head-Related Transfer Functions (HRTF). In their research on HRTF modelling, Brown et al. (1998) suggested that listening in humans depends on three factors:

1. Head shadow and Interaural Time Difference

2. Shoulder Echo

3. Pinna Reflections

Head shadow modelling involves the approximation that the torso is affected by a sound wave such that the head shadow coefficients can be defined as:

$$s_z = \frac{-2w_0}{\alpha(\theta)} \tag{5}$$

$$s_p = 2w_0 \tag{6}$$

$$w_0 = \frac{c}{a} \tag{7}$$

Here, α denotes the radius of the head, c is the speed of sound and w_0 is the angular frequency of the sound source diffusing around the head.

Once the head shadow coefficients have been derived, an integral sample delay is implemented that delays a signal $x(n)$ by a whole number of samples. To achieve efficient recursive cancellation between the ipsilateral and contralateral ears, a fractional delay needs to be derived. Given the integral delay values, two points f and $(1-f)$ are linearly interpolated between two whole number delay values. Thus, given the radius of the head and the speed of sound, the difference in the arrival of sound to both the ears can be calculated accurately. The process of recursive cancellation can now be implemented by doing the following steps:

1. Calculating the time difference values between both ears for the left and right channel

2. Cancellation is achieved by phase inverting and delaying (in samples) the intended crosstalk signal for both channels

3. A threshold is set above which the signal cancellation takes place

4 Comparison

After implementing the two methods, a sample binaural audio file was processed through the two methods. The following are the spectrograms of the left channel of the resulting crosstalk cancellation processed signals:

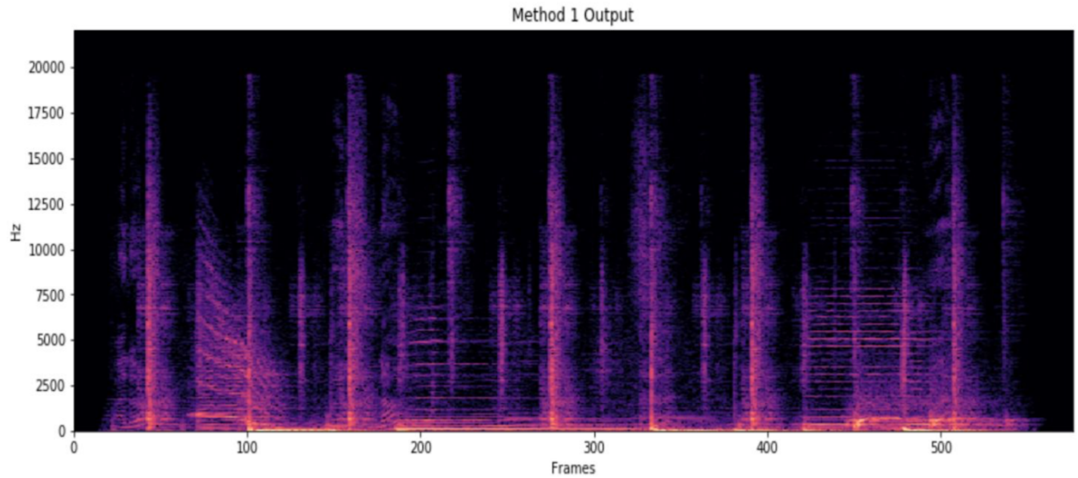

Figure 3: Method 1 Spectrogram

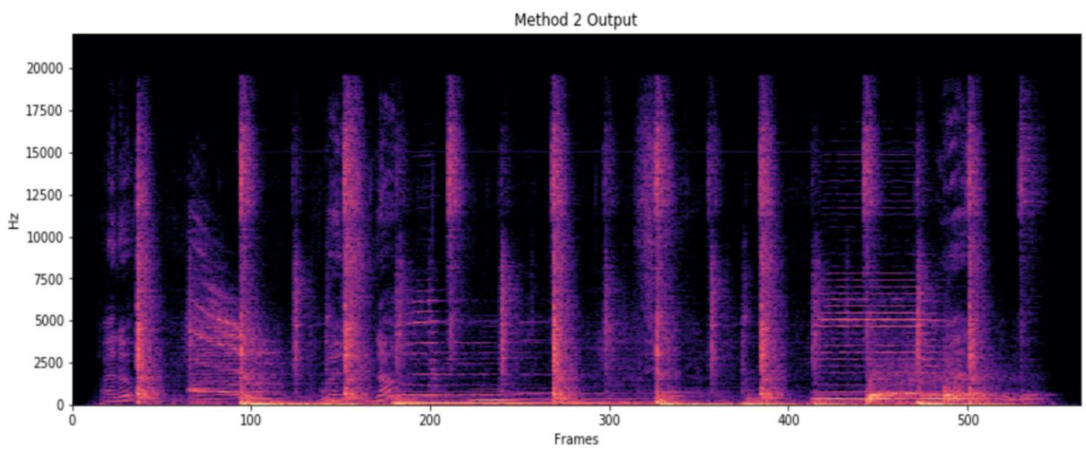

Figure 4: Method 2 Spectrogram

From the above spectrograms, it can be seen that the crosstalk cancellation method that makes use of a regularization factor β , the peaks in high frequency content is lower. This implies an alleviation of the tonal distortion around the high frequency regions. However, an interesting observation was recorded from listening to the two methods. Upon listening to the files, the recursive crosstalk cancellation method was deemed to be wider.

References

Brown, C., Duda. R. (1998). "A structural model for binaural sound synthesis ." *IEEE Trans. Speech and Audio Process., 6(5)*, pp. 476 - 488.

Roginska, A., & Geluso, P. (Eds.). (2017). *Immersive Sound: The Art and Science of Binaural and Multichannel Audio.* Taylor & Francis.

Zolzer, Udo. (2002). "DAFX - Digital Audio Effects." *Edited by Udo Zolzer. Chichester, England: John Wiley & Sons,* pp. 575 - 585