3D Audio Final Project

Implementation of a Frequency-dependent Crosstalk Cancellation System

Devyani Hebbar

Course: MPATE-GE 2613 3D Audio

Instructor: Dr. Agnieszka Roginska

1. **Abstract**

This study explores the implementation of Frequency Dependent Crosstalk Cancellation (XTC) using the Matrix Inversion method in Python. A simple non-frequency dependent system is implemented. This system is then improved by using frequency dependent regularization both with and without branch based regularization. It is observed that the branch based regularization approach outperforms both the former approaches by resulting in a more balanced frequency response with better attenuation levels.

2. **Background**

In the reproduction of stereo and binaural audio over headphones, the signals from each channel only reach that ear of the listener that they were meant to reach. However, in the reproduction of such audio over loudspeakers, the signals from each channel also travel to the opposite ears. This phenomenon is known as crosstalk. Crosstalk is especially an issue in binaural audio because it introduces ambiguity in sound source localization. A binaural signal is encoded with spatial cues using Head Related Transfer Functions (HRTFs) that are designed to mimic how sound interacts with the listener’s head, torso, and ears to create the perception of direction and distance. Crosstalk disrupts these spatial cues by allowing signals intended for one ear to reach the other, distorting the precise localization information encoded by the HRTFs. The purpose of a crosstalk cancellation system is to remove unwanted signals and resolve ambiguity in sound source localization *(Roginska & Geluso, 2017).*

3. **Motivation**

Crosstalk cancellation (XTC) is essential for creating accurate 3D sound through loudspeakers, especially in areas like virtual reality and advanced audio systems that demand high spatial precision. Without effective XTC, sound from one loudspeaker reaches the opposite ear, disrupting the spatial cues in binaural audio that help listeners identify where sounds are coming from. However, the main challenge is not just to remove this crosstalk, but also to keep the original sound's quality intact. Most XTC systems cause spectral coloration, which changes the sound’s natural characteristics and limits the practical use of XTC.

Frequency-dependent XTC offers a way to address spectral coloration while ensuring effective cancellation. Unlike non-frequency-dependent methods that reduce sound indiscriminately, this approach adjusts attenuation based on the audio spectrum, avoiding extreme distortions. However, it can still create problems like over-attenuation in certain frequency ranges. To overcome this, Edgar Choueiri (2010) proposed branch-based frequency-dependent regularization. This method divides the frequency spectrum into branches and applies specific adjustments to each band, flattening the response while reducing coloration.

4. Methodology

4.1. Implementation of Non-frequency Dependent Crosstalk Cancellation

HRTFs for the four signals involved in this system were obtained. The names of the HRTFs and their corresponding signals are given below:

H_LL: Left loudspeaker to left ear

H_LR: Right loudspeaker to Left ear

H_RL: Left loudspeaker to Right ear

H_RR: Right loudspeaker to right ear

In this case, I considered a stereo speaker setup where the left speaker is placed at 330 degree azimuth and right speaker is placed at 30 degree azimuth. I used the MIT zero elevation dataset to obtain these HRTFs.

The FFT was applied to the HRTFs to transform them from the time domain to the frequency domain. Once this was done, they organized into a 2x2 matrix H at each frequency bin, representing the transfer from each loudspeaker to each ear. A small regularization parameter delta ($\boldsymbol{\delta}$) was added to the diagonal elements to improve numerical stability and avoid inverting nearly singular matrices. The matrix H was then inverted by taking the reciprocal of its determinant and multiplying it with its adjoint, resulting in H^{-1}.

The loudspeaker signals S_L and S_R, i.e., the signals to be sent to each channel of the loudspeaker were calculated using the equation *(Choueiri, 2010, p. 5):*

$$S = H^{-1} \cdot D$$

Where S is the matrix containing S_L and S_R, and D is the matrix containing the desired signals L_desired and R_desired, which are simply the left and right channels of the input binaural signal respectively.

The ear signals (signals reaching either ear of the listener) before crosstalk cancellation were calculated as *(Choueiri, 2010, p. 4):*

$$Ear\_L\_before = H\_LL * L\_desired + H\_LR * R\_desired$$

$$Ear\_R\_before = H\_RL * L\_desired + H\_RR * R\_desired$$

Similarly, The ear signals after crosstalk cancellation were calculated as *(Choueiri, 2010, p. 5):*

$$Ear\_L\_after = H\_LL * S\_L + H\_LR * S\_R$$

$$Ear\_R\_after = H\_RL * S\_L + H\_RR * S\_R$$

The crosstalk levels before and after putting the signal through the system were calculated using the equations below, and the total crosstalk attenuation level was calculated by computing their difference.

$$Crosstalk\_L = Ear\_L - L\_desired$$

$$Crosstalk\_R = Ear\_R - R\_desired$$

4.2. Implementation of Frequency Dependent Crosstalk Cancellation

There are certain frequencies that get boosted by H^{-1}, where the non-frequency-dependent system's envelope spectrum exceeds the desired intensity level. This causes the system to have an uneven frequency response and causes the output signal to sound extremely lowpass filtered after crosstalk cancellation *(Choueiri, 2010, p. 12).*

In order to attenuate these boosted frequencies, a frequency dependent regularization parameter is used. The matrix $H^H \cdot H$ was calculated, and then split into three components using the Singular Value Decomposition method. This gives us the diagonal component that contains $\sigma_{min}(f)$, the smallest singular value of H at frequency f. The frequency dependent regularization parameter was then calculated as:

$$\lambda = \gamma * \sigma_{min}(f)$$

The new H^{-1} matrix was then calculated as:

$$H_{reg}^{-1} = (H^H H + \lambda I)^{-1} H^H$$

This regularized matrix ensures better conditioning and prevents over-amplification at frequencies where H is nearly singular.

The loudspeaker signals S_L and S_R were calculated as in the previous section, and therefore, total crosstalk attenuation was calculated in the same manner.

4.2.2. Branch-based Frequency Dependent Crosstalk cancellation:

According to Choueiri (2010), using a general frequency dependent regularization parameter is not enough, since different bands of the frequency spectrum are coloured in different ways. He therefore proposes a branch based system, where each branch has its own regularization parameter. The three branches are as follows:

- Branch I: For frequencies where the out-of-phase response dominates, a specific regularization parameter λ_1 is applied.

- Branch II: For frequencies where the in-phase response dominates, a different regularization parameter λ_2 is applied.
- Branch P: For frequencies where the filter is already perfect (no regularization needed).

The new γ was calculated as

$$\gamma = 10^{\gamma db/20}$$

Where γdb is the desired level of attenuation or the target envelope intensity for the frequency response of the crosstalk cancellation filter. This determines the threshold for how much the regularization parameter should modify the filter to avoid artifacts while maintaining effective cancellation.

λ_1 and λ_2 are calculated according to the following equations *(Choueiri, 2010, p. 12):*

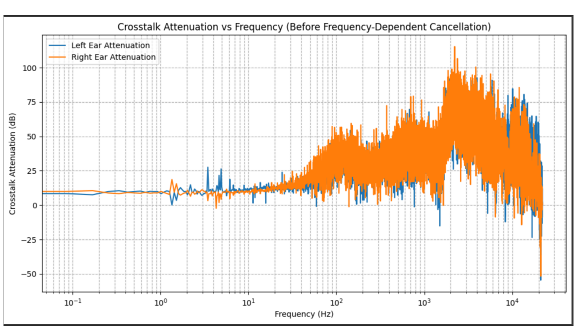
$$\lambda_1(\omega c) = \sqrt{}\,(g^2 + 2g\cos(\omega c) + 1 \,/\, g^2 - 2g\cos(\omega c) + 1) - 1$$

$$\lambda_1(\omega c) = \sqrt{}\,(g^2 - 2g\cos(\omega c) + 1 \,/\, g^2 + 2g\cos(\omega c) + 1) - 1$$

Where g is a constant.

The matrix H_{reg}^{-1} was then calculated in the same manner as in the previous subsection, with the value of λ changing according to the corresponding frequency branch.

5. Results



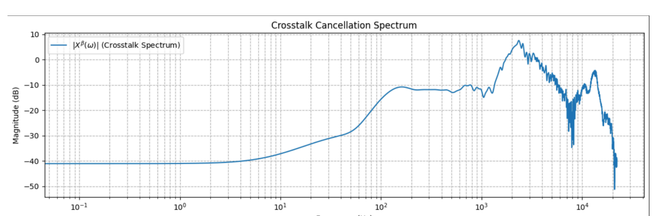


Fig.1 (a)

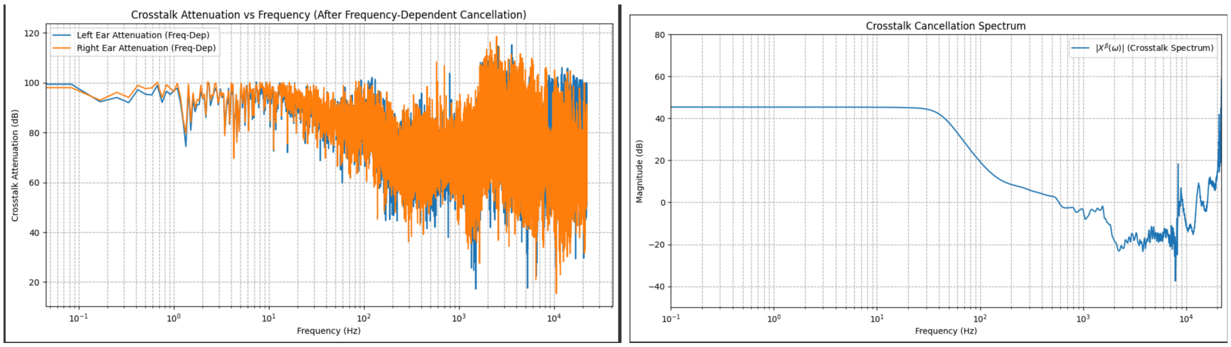


Fig. 1 (b)

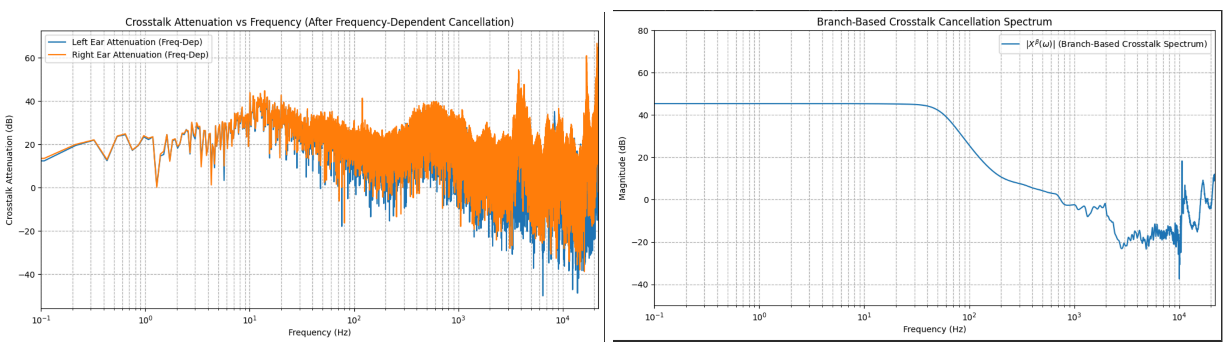


Fig. 1 (c)

Fig.1 a: Frequency vs Crosstalk Attenuation when applied to a binaural signal (left) and Frequency Response Spectrum (right) of the non-frequency-dependent system
Fig.1 b: Frequency vs Crosstalk Attenuation when applied to a binaural signal (left) and Frequency Response Spectrum (right) for the frequency-dependent system (without branch based regularization)
Fig.1 c: Frequency vs Crosstalk Attenuation when applied to a binaural signal (left) and Frequency Response Spectrum (right) for the frequency-dependent system (with branch based regularization)

The images on the left show how much each frequency is attenuated after each crosstalk cancellation system has been applied to the same binaural signal. The binaural signal used here is a recording that was done using the Neumann KU-100 Head at Bobst Library NYU. The signal is rich in terms of spatial cues (sounds coming from various angles relative to the listener) and also has a wide range of frequencies in its spectrum. The images on the right show the general behaviour of each crosstalk cancellation system (without applying the system to a specific signal).

As seen in Fig. 1 (a), the non-frequency dependent system has low attenuation in the low frequency range, and extremely high attenuation in the middle and high frequency ranges. Therefore, it exhibits low pass filtering as expected. Fig. 1 (b) shows that the frequency dependent system that does not use branch based regularization causes a spike in attenuation of low frequencies. While the attenuation is more consistent across the spectrum as compared to the non-frequency dependent system, the attenuation levels are extremely high (20 to 120 db), which severely colours the input signal, stripping it of its intensity across the entire spectrum. Fig. 1 (c) shows the frequency dependent system that uses branch based regularization. While its response is not perfect, it shows significantly higher balance across the frequency spectrum as compared to the first system, and has significantly lower attenuation levels (10 to 50 db) as compared to the second system, outperforming both of them.

6. Conclusions and Future Work

This study showed that using the Matrix Inversion method is effective in reducing crosstalk in spatial audio over loudspeakers. It also confirmed that this method induces heavy spectral colouration. The frequency-dependent crosstalk cancellation system with branch-based regularization demonstrates the most balanced performance, minimizing artifacts while maintaining consistent attenuation across the frequency spectrum. Future work could explore refining the branch-based approach by incorporating adaptive regularization techniques that dynamically adjust to signal content. One approach to this could be using Machine Learning techniques. Additionally, testing with diverse input signals and various values for the different constant parameters used could also help optimize the system.

7. References

1. Roginska, A., Geluso, P. (2017). *Immersive Sound: The Art and Science of Binaural and Multi-Channel Audio.*

2. Choueiri, E. (2010). *Optimal Crosstalk Cancellation for Binaural Audio with Two Loudspeakers*. Princeton University Journal, Volume 28.