Beamforming Based Speech Enhancement and Noise Suppression

Autumn Coe, Hilary Mogul, Yiting Zhang | ECE 472 Audio Signal Processing

Abstract

Speech enhancement is used in all kinds of applications including hearing aids, mobile phones, speech recognition, etc. There are lots of DSP algorithms for speech enhancement and noise cancellation, including filtering techniques, spectral restoration etc. This project focuses on superdirective beamforming for speech enhancement in a hearing aid application. Superdirective beamforming is advantageous in a hearing aid situation because it isolates a source from a specific direction, while still maintaining some semblance of directionality. Wiener postfiltering is then also used to reduce the SNR of the signal.

Objectives

- Enhance a single source in a noisy environment based on multiple inputs
- Implementing findings from [1] in a real time/VST plugin setting
  - Dual input-output superdirective beamforming with adaptive postfiltering to further suppress noise

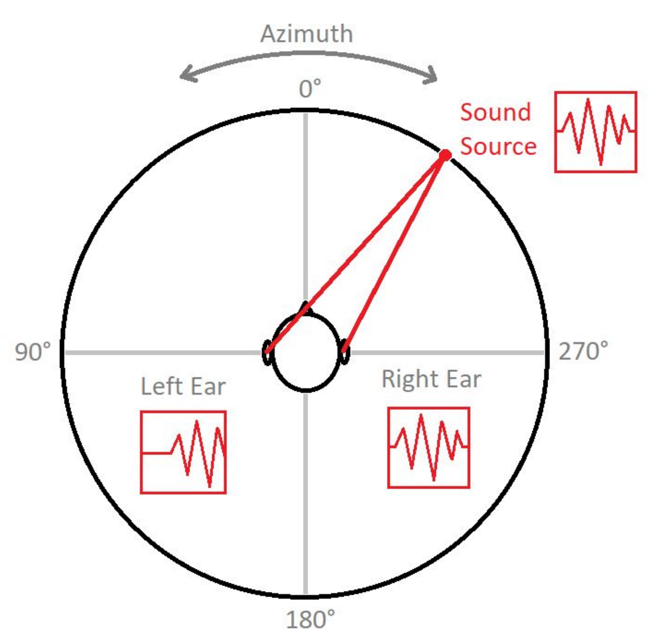

Figure 1: Dual channel binaural input from a mono source

Overview and Background

- Beamforming
  - Given an array of M microphones with known distances apart, and a known steering angle to a source, combine their signals somehow to generate a mono signal of the source
- Superdirective Binaural Beamforming
  - Using the beamformed signal and the original M inputs, regenerate M outputs with the signal isolated from their original steering angle.

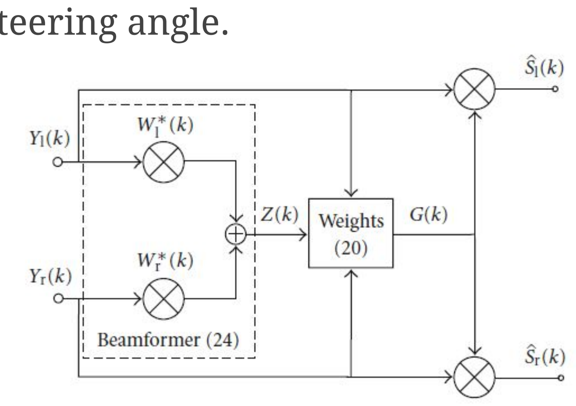

Figure 2: Superdirective binaural input-output beamformer. [1]

- Wiener Filter
  - Objective: Generate filter coefficients to approximate an unknown signal using a noisy one, assuming some knowledge of their cross correlation.

Mathematical Derivation

Originally, in [1], the system is broken down into three components:

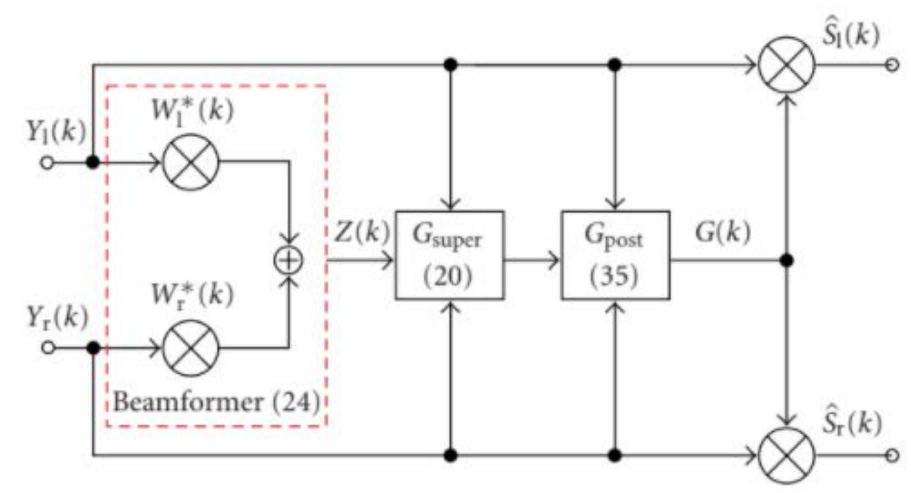

Figure 3: Superdirective input-output beamformer with postfiltering. [1]

Using the following equations to calculate G(k), the filter applied to the binaural signal, where $D_{L/R}(\theta,k)$ represents the HRTF of some steering angle:

$$W_{L/R}(\theta_S,k) = \frac{\Phi^{-1}_{MM}(k)D_{L/R}(\theta_S,k)}{D^{-1}_{L/R}(\theta_S,k)\Phi^{-1}_{MM}(k)D_{L/R}(\theta_S,k)}$$

$$G(k) = G_{super}(k)G_{post}(k)$$

$$G_{super}(k) = \frac{Z(k)}{|Y_L(k)|+|Y_R(k)|}$$

$$\alpha^{phy}_{L/R}(\theta_S,k) = |D_{L/R}(\theta_S,k)|$$

$$G_{post} = \frac{|Z(k)|^2}{|Y_L(k)|^2+|Y_R(k)|^2}\frac{\alpha^{phy}_L(\theta_S,k)^2+\alpha^{phy}_R(\theta_S,k)^2}{(\alpha^{phy}_L(\theta_S,k)+\alpha^{phy}_R(\theta_S,k))^2}$$

Since we have precalculated HRTFs, we can factor the portion of this system after the beamformer into two components:

$$G_{adaptive}(k) = \frac{|Z(k)|^3}{(|Y_L(k)|^2+|Y_R(k)|^2)(|Y_L(k)|+|Y_R(k)|)}$$

$$G_{static}(\theta_S,k) = \frac{\alpha^{phy}_L(\theta_S,k)^2+\alpha^{phy}_R(\theta_S,k)^2}{(\alpha^{phy}_L(\theta_S,k)+\alpha^{phy}_R(\theta_S,k))^2}$$

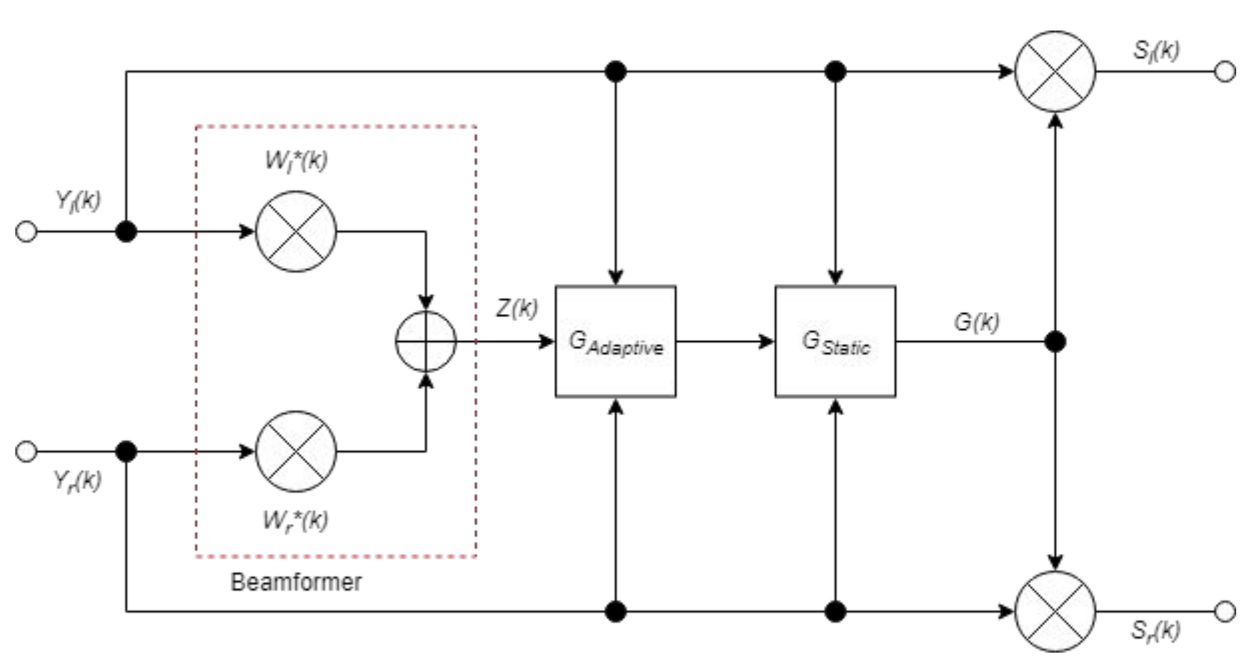

Figure 4: Superdirective input-output beamformer with postfiltering factored into adaptive and static phases.

We can calculate the static portions in MATLAB based on the HRTFs, outputting .wav files of the time domain impulse responses to be loaded dynamically into a JUCE-based VST plugin.

Implementation Details

The system itself was implemented as a plugin built in JUCE, an audio application framework with DSP capabilities. We load these .wav files and load into Convolution processor chains for the static filters. We then use the static beamformer to calculate z[n], and perform an FFT on the signal to obtain Z(k), which is used to calculate $G_{adaptive}(k)$. We then perform an IFFT to get the time domain form of this signal, and load it into a processor chain to convolve the original inputs with the cascade of $G_{adaptive}(k)$ and $G_{static}(\theta,k)$. The GUI for this system provides a rotary slider to choose which angle, and a means to select a base folder to load the static impulse responses from.
To simulate hearing aids, the Roland Binaural in-ear monitors are used. These are earbuds equipped with microphones used for binaural recording.

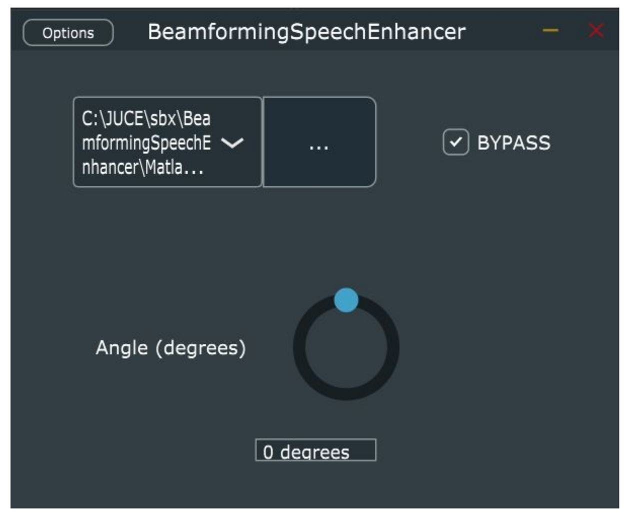

Figure 5: User interface of plug-in

Figure 6: Hardware used

Results

This was subjectively tested by placing white noise generators around a room, and having one user wear the headphones while another speak towards the former at quieter and quieter levels. The user wearing the headphones was able to continuously hear the speaker, and then compare the same levels while not wearing the headphones, which was found to be unintelligible. Thus, the system is successful in directionally enhancing a signal, and suppressing the noise around it.

Future Work

- Calculate the HRTFs instead of using preloaded options to enhance the angular resolution.
- Include some sort of visualization in the plugin of the signals before and after the system is applied.

Citation

[1] T. Lotter and P. Vary, "Dual-Channel Speech Enhancement by Superdirective Beamforming," *EURASIP Journal on Advances in Signal Processing*, vol. 2006, no. 1, 2006.

[2] "JUCE," *JUCE*. [Online]. Available: https://juce.com/. [Accessed: 28-Apr-2018].