Overview

This document details the design and configuration of the ideal binary mask, which was implemented for our Binary DST study. Since a new study (Binary Mask for Children) plans to use this ideal binary mask, now is a good time to review the design and configuration of this implementation.

The design is similar to that used by Brungart et. al. (2006). However, there are some differences. These differences between our implementation and Brungart's are described in this document.

Gammatone Filter

We used an implementation of the gammatone filter from Ning Ma. Details on this implementation can be found here.

Gammatone Filter Bank

Theoretically, the formula for the bandwidth of the gammatone filter is the following (from Ma):

$BW_{gammatone} = 1.019*ERB$

$ERB = 24.7 + 0.108*f_c$ where f_c is the center frequency

The filters in our gammatone filter bank are spaced such that they overlap at their 3 dB points.

For example, Figure 1 below shows the frequency response of two consecutive filters in our filter bank. (The frequency response was generated by passing Gaussian white noise through our filter bank and plotting the spectra of the output of two consecutive filters.)

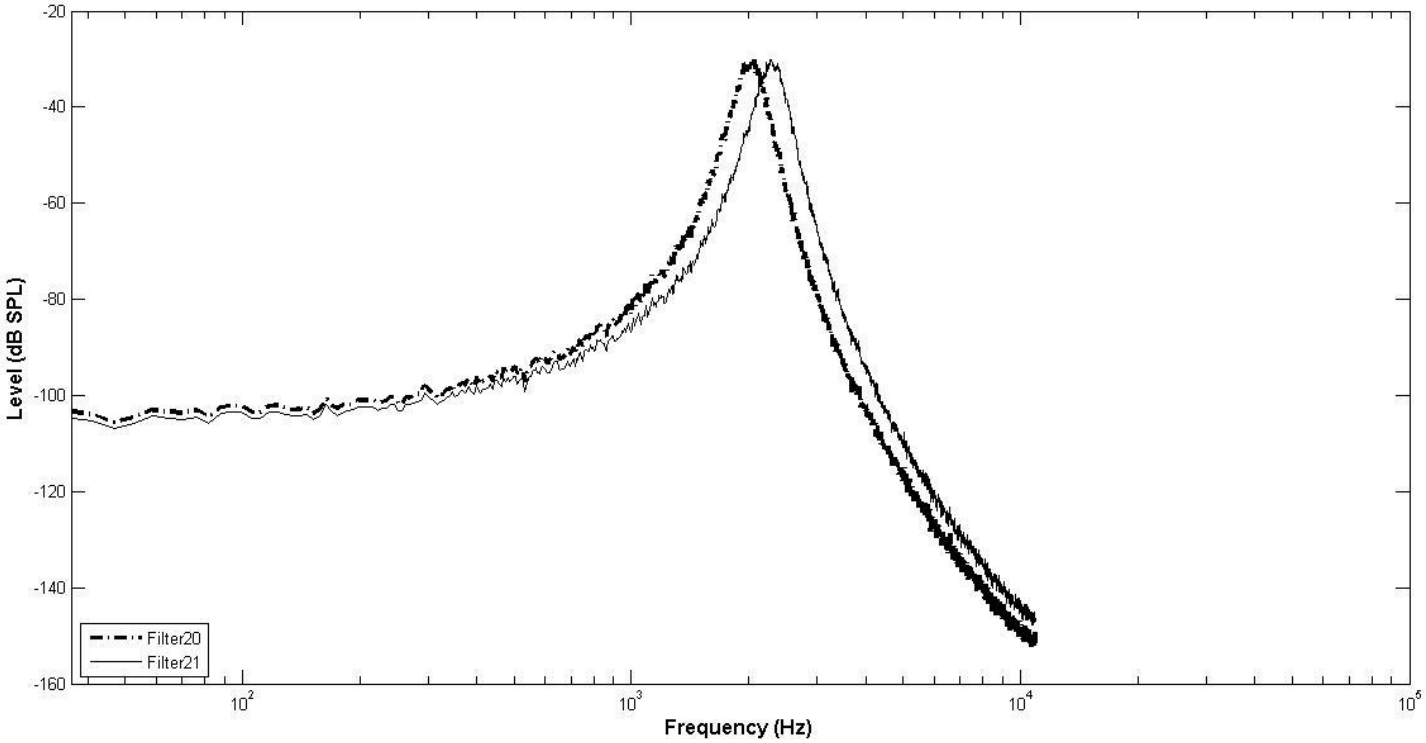

Figure 1: Magnitude of filter response from two consecutive gammatone filters

If you space the gammatone filters such that they overlap at their 3 dB points, you need only 35 filters to span the range of center frequencies from 50 Hz to over 11 kHz. This is illustrated by Figure 2 below.

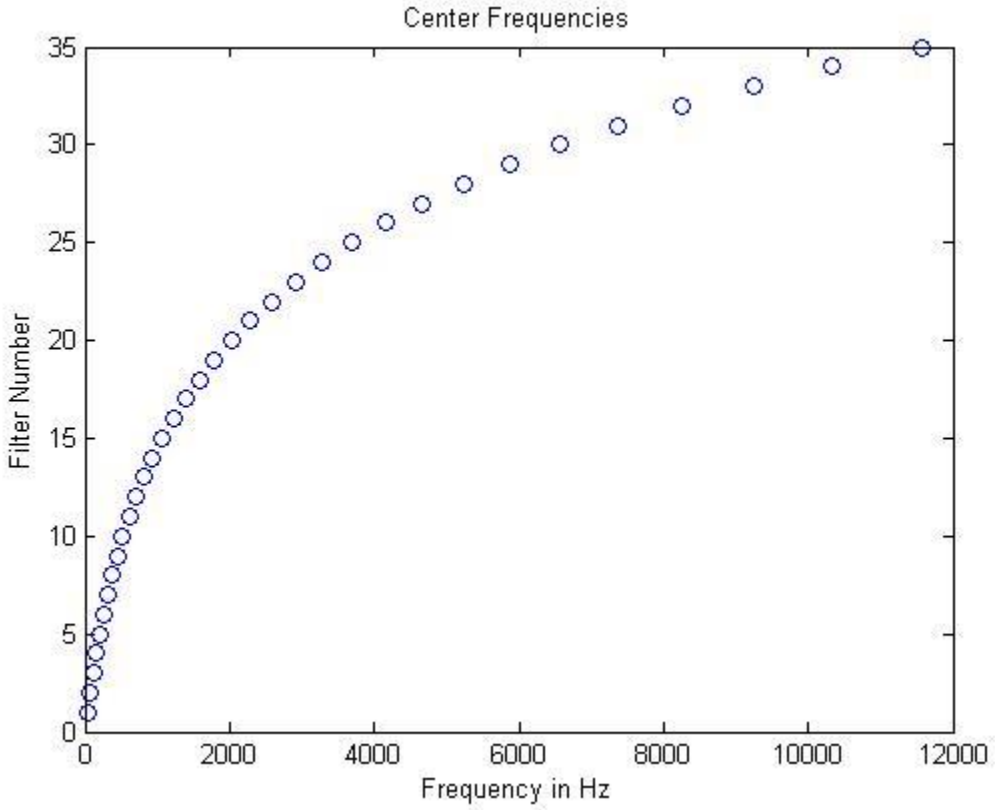

Figure 2: Center frequencies of the gammatone filters in our filter bank

The following table details the differences between our filter bank configuration and the one used by Brungart et. al.

Parameter	**Brungart's Setting**	**Our Setting**	**Comments**
Lowest Center Frequency (Hz)	80	50	
Highest Center Frequency (Hz)	5000	11569	
Number of Filters	128	35	Our filters overlap at 3-dB points.

Binary Mask

To create the binary mask, we must "chop" both the signal and masker into time-frequency (t-f) bins. Within these t-f bins, we compare the SNR to a user-defined threshold (e.g. -6 dB) to see if the mixture (signal plus noise) is allowed to pass through that t-f bin. Specifically, the following steps are performed per trial:

1. Scale the signal and the noise (masker), such that
    a. The level of the mixture (before filtering) is 75 dB SPL.
    b. The SNR (before filtering) is a specified value, as determined by the adaptive track.
2. Chop up the scaled signal and scaled masker into frequency bins; in other words, filter as follows:
    a. Pass the scaled signal through the gammatone filter bank. This gives us 35 filtered (signal) time series.
    b. Pass the scaled masker through the gammatone filter bank. This gives us 35 filtered (masker) time series.
3. Chop up the filtered signal and filtered masker into time bins.
    a. For each filtered time series, create 50%-overlapping segments of duration 20 ms.
    b. A periodic Hann window is applied to each segment. The overlapping segments satisfy constant overlap-add requirements (COLA).
4. For each t-f bin, calculate the SNR. Compare this SNR to the user-defined SNR Threshold. If the SNR is greater than the SNR Threshold, we consider this time-frequency bin "enabled". Figure 3 below illustrates one instance of the binary mask, showing which bins are enabled.
5. If the t-f bin is enabled, its output is the mixture for that t-f bin (the filtered signal plus the filtered masker for that filter and that windowed time segment); otherwise the t-f bin's output is set to zero for all samples in that bin.

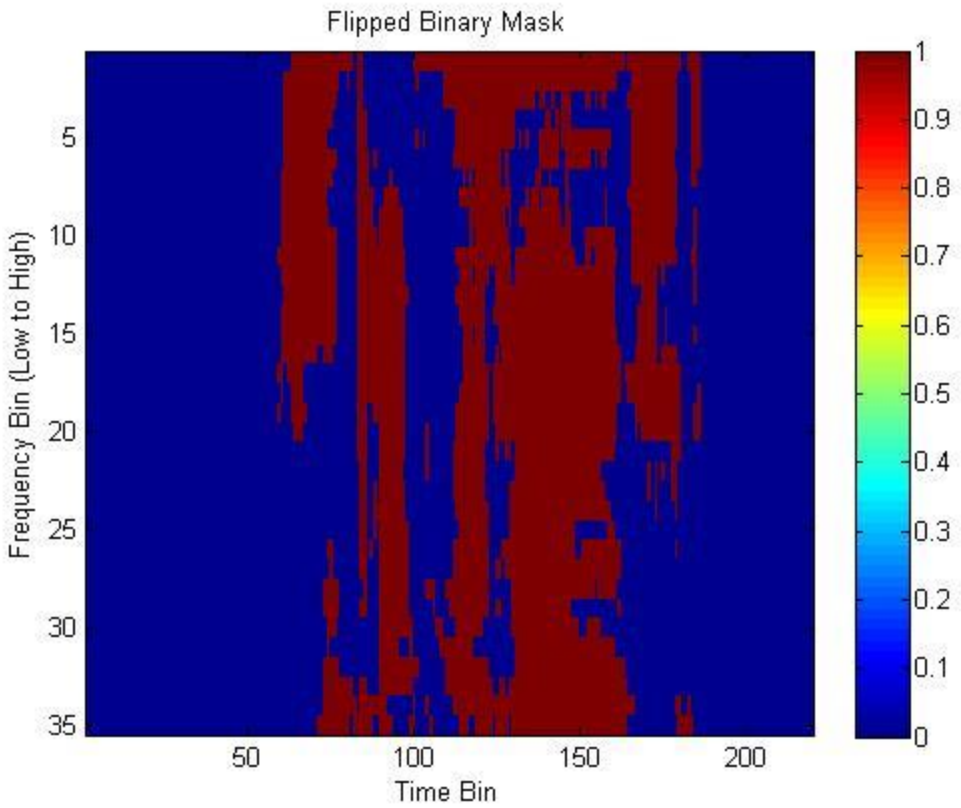

Figure 3: Sample binary mask, where enabled bins have a value of 1 (red), and disabled bins have a value of 0 (blue)

Reconstruction of the Stimulus

To reconstruct the stimulus, all we need to do is to sum the outputs of all the time-frequency bins of the binary mask. This reconstructed stimulus is then presented to the subject.

If the gammatone filter bank introduces phase distortion, one may try "reverse filtering" as a step in the reconstruction process. It should undo the phase distortion. In our case, phase distortion was low, so reverse filtering was not needed. (See Appendix for plots on phase comparisons.)

Note that Brungart *did* use reverse filtering. It's also important to note that Brungart used many more filters than we did. By overlapping the filters at their 3-dB points, we have reduced phase distortion.

Note that the level of the mixture is 75 dB *before* it is passed into the filter bank. As the mixture's SNR gets lower, less of the mixture will get through the binary mask. So, if the adaptive track sets the SNR lower, the stimulus will get weaker.

Appendix

The following plots were generated to evaluate the quality of the reconstruction. In this case, there is no masker. A signal is passed through the filter bank and then reconstructed without applying a binary mask. This allows us to measure the phase distortion from the gammatone filter bank. The signal is the word “birthday”.

Compare the Phase of the Spectra

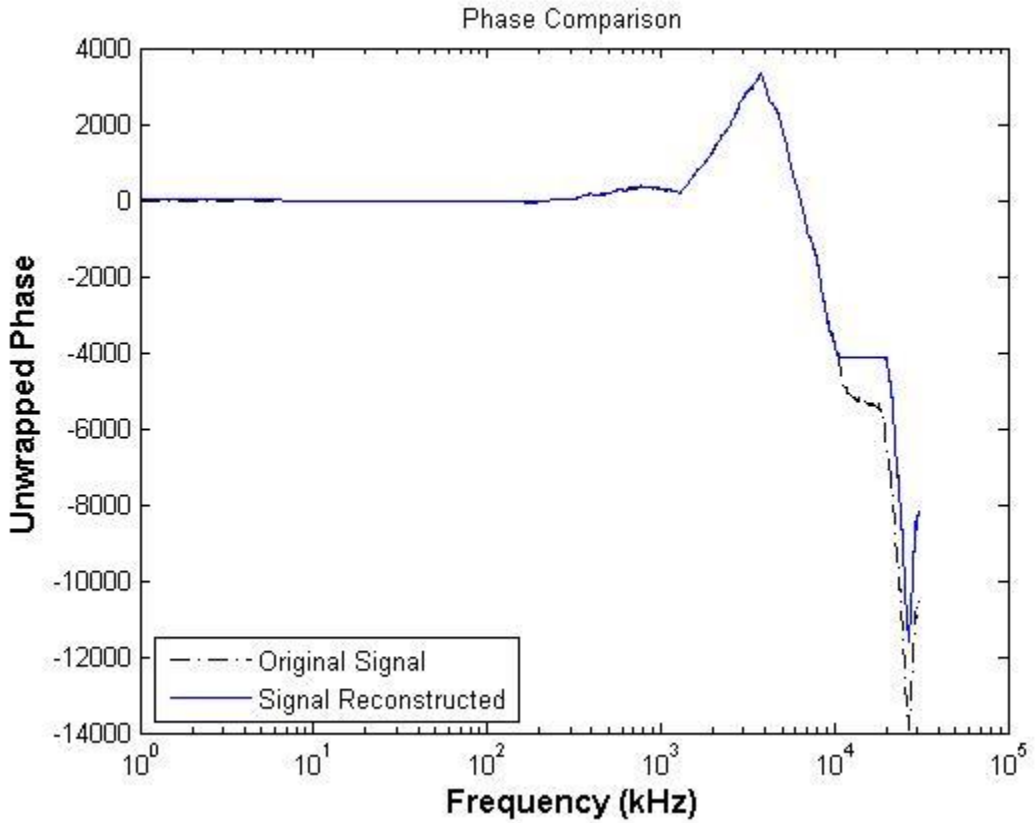

Figure 4: Phase of spectrum--reconstructed stimulus compared to original stimulus

Compare the Magnitude of the Spectra

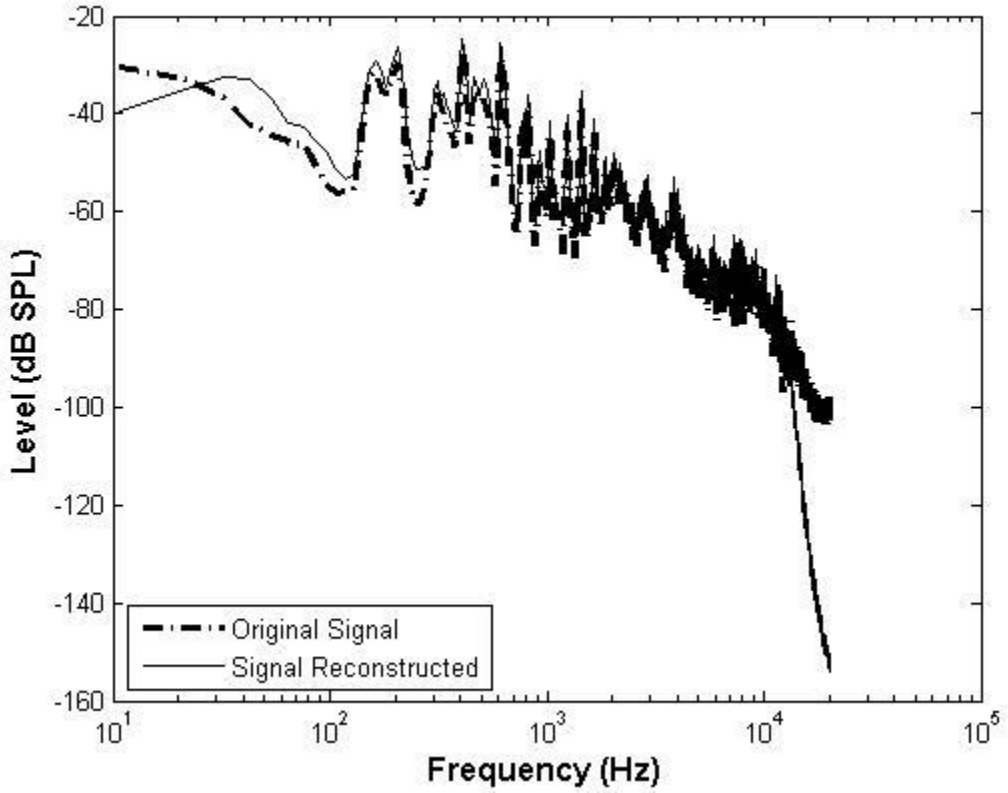

Figure 5: Magnitude of spectrum--reconstructed stimulus compared to the actual stimulus

Differences Between Our Approach and Brungart's

The differences between our approach and Brungart's have been discussed above, but they are summarized here:

1. The configuration of the filter bank is different, as discussed in "Gammatone Filter Bank".
2. Consequently, reverse filtering is not used during reconstruction, as discussed in "Reconstruction of the Stimulus".

References

Brungart, D.S., Chang, P.S., Simpson, B.D., and DeLiang, W. (2006). "Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation", J. Acoust. Soc. Am. 120, 4007-4018.