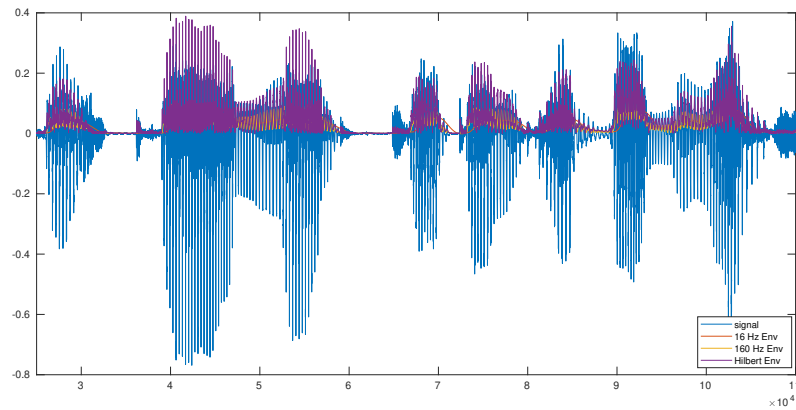


## 1 Part A | Introduction to cochlear-implant signal processing

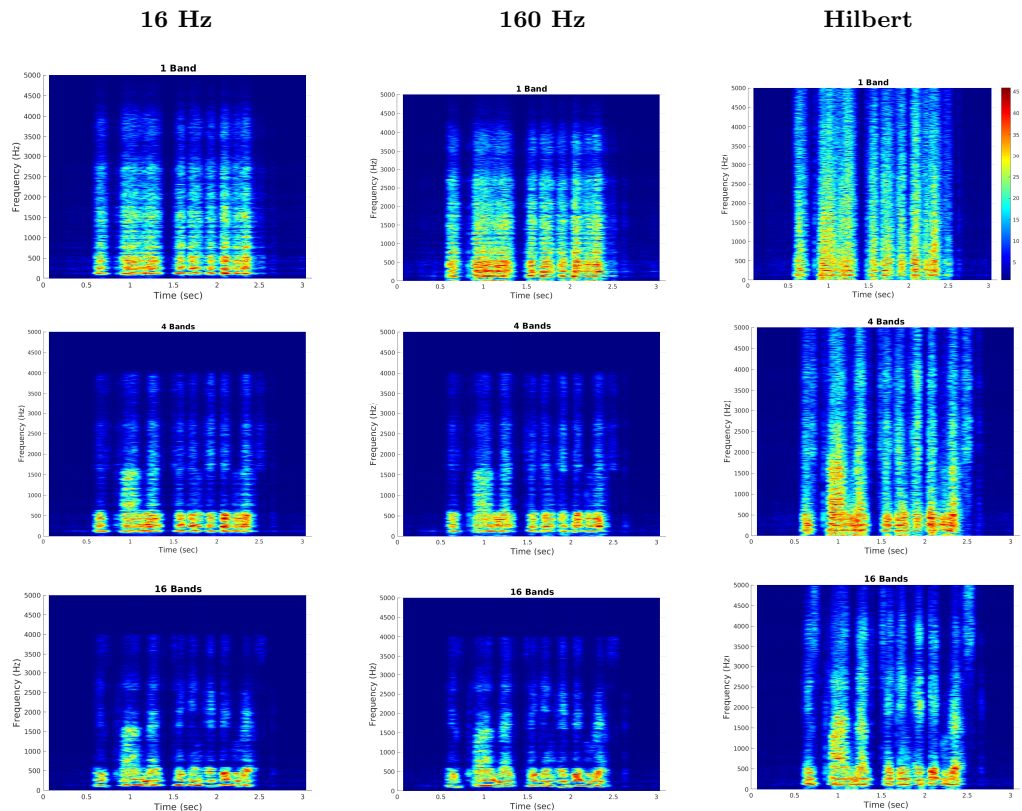
- a) Filter-bank reconstructions were generated using the `quad_filt_bank.m` function from Delgutte within the `get_Stimuli.m` function I created. The bands were set using the `equal_xbm_bands.m` function from Delgutte. This function spaced the frequency bands to match the frequency distribution of the basilar membrane using Liberman's cat cochlear frequency map, scaling it to match the ~20 kHz hearing range of humans. For cochlear implant user simulation, this is important, since we assume cochlear implant electrodes (bands) are equally spaced. However, the frequency-specific IHCs in the cochlea are not equally spaced from 0-4 kHz, so this correction helps.

Refer to B in my `get_Stimuli.m` code for the filter coefficients.

- b) Below are the different envelope extractions visualized for the 1 band filter-bank. It is apparent that the Hilbert envelope is a more liberal than the 16 and 160 Hz envelope extraction methods, allowing it to capture a more detailed picture of amplitude variations in the sound being played.



- c) Here are some spectrograms (bandwidth 30 Hz, dynamic range 45 dB) that demonstrate proof of a working filter-bank:

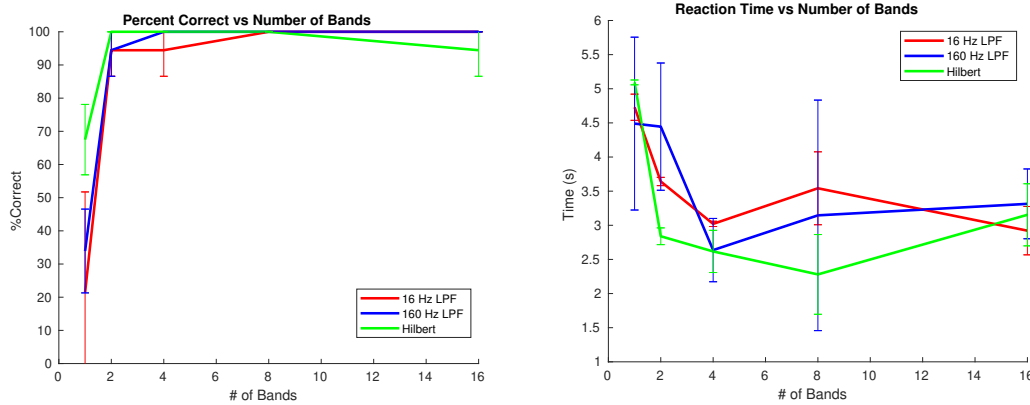


The sharper cutoff at 4 kHz in the 16 and 160 Hz envelope extraction can be explained by the 4 kHz LPF filter I used to emulate the Shannon et. al. (1995) study. Additionally, it is (again) clear that the Hilbert envelope extraction and presentation with a noise carrier captures more detail than the 16 and 160 Hz envelope extractions. All spectrograms tend to have closer representation of the raw recording as the number of bands increase. *I apologize for these figures not being evenly sized. It's an eyesore but I had to get this done.*

- d) The way we tested this was by taking ten of the SPIN-R sentences, and creating 16/160 Hz and Hilbert envelope extractions & noise carrier signal for all ten sentences. Our code then randomly presented these stimuli and asked the user to guess which sentence was presented. We presented a total of 120 sentences/subject. We tracked correctness, in addition to reaction time. The code for this may be found in `runTrialA.m`.

From our results, it is clear that the Hilbert envelope allows for best representation of the sentence most-noticeably at the 1-2 Band filter-bank level. Above that, the 16/160 Hz envelope extraction works fairly well. It is important to consider the fact that our study was more primitive than the Shannon study. Instead of counting the number of words typed by the user, we went for a more simplistic approach to save time (10 alternative forced choice). Additionally, one of our subjects transcribed the sentences and de-bugged the code, so was rather familiar with the 10 sentences that were used, which results in better accuracy with sentence detection. **COMMENT ON 16 Hz vs others after Joe's results and mention our findings are similar to shannon!!!!**

Though our reaction time results are noisier and **don't demonstrate much variability between conditions**, they may demonstrate an increase in confidence of sentence identification as the number of bands increase.



## 2 Part B | Exploring potential for frequency modulation information to improve CI signal processing

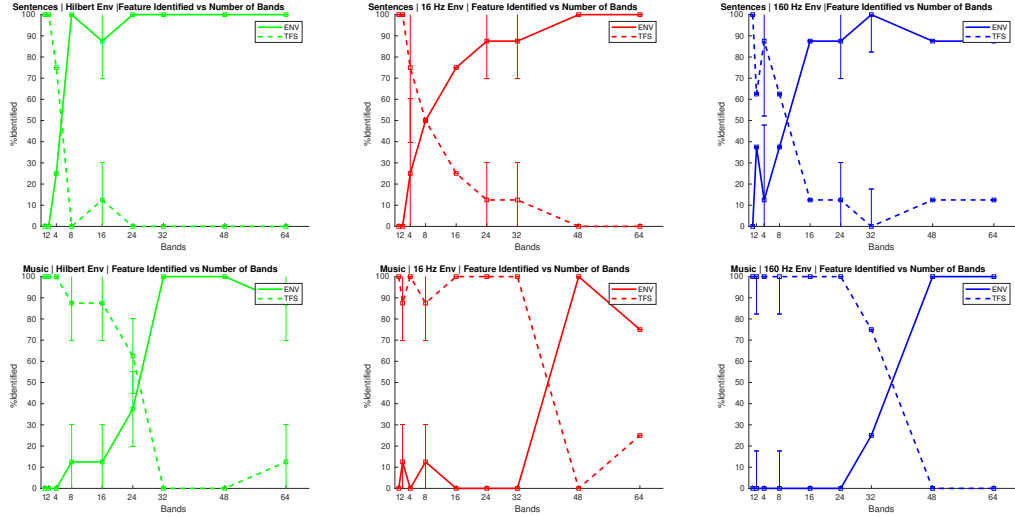
- a) The Delgutte implementation of acquiring the analytic signal is a smart approach because it prevents the computation of multiple large FFTs. Typically, the way the `hilbert` function in Matlab works is by using a one-sided discrete-time frequency-domain approach outlined in a 1999 paper by Marple. While this method works well for obtaining an analytic signal, it would be computationally expensive in the chimera application because the code would have to first apply our band-pass filters to the signal, then send multiple time-domain signals to `hilbert` which would do an FFT and inverse-FFT to return the analytic signal.

The Delgutte code combines a few steps into one very efficient method. The `quad_filter_bank` function generates complex FIR filters *in quadrature*. This means the filter coefficients from `b` are in terms of *cos* and *sin* and where the *cos* part is real and the *sin* part is imaginary (and represents the Hilbert Transform). Since the analytic signal is in this form (since the Hilbert transform is a change in phase by  $+\frac{\pi}{2}$  for neg. frequencies and  $-\frac{\pi}{2}$  for pos. frequencies), the envelope and TFS information can be easily extracted by calculating the magnitude and phase angle of the filtered output, respectively (without having to use `hilbert`). The chimera is then synthesized by multiplying the envelope from one signal by the TFS of another. The real part of `b` simply represents the filter-bank bandpass filters, and can be used to refilter the chimera to compensate for filtering-related phase shifts— as Delgutte does.

To create the noise carrier to make speech-noise chimeras, Delgutte uses *spectrally-matched* noise. This means the noise carrier has the same frequency content as the speech signal. The way this is

implemented is by taking an FFT of the speech signal, randomly changing the phase using  $e^{2\pi j\omega_{random}}$ , and taking the inverse FFT.

- b) Our perceptual task consisted of asking the subject which sentence/music they heard most clear, using a single pair of chimeras (S1/S2, M1/M2). The code, found in `runTrialB_music` or `runTrialB_sentence`, noted if the subject chose the envelope or TFS. This was a 2 alternative forced choice study. Therefore, the TFS and ENV curves presented for each band are exactly inverse, unlike the Smith study, which asked subjects to identify multiple melodies if they heard them. The results are shown below (in green):



- c) Using 16 Hz and 160 Hz envelope extractions resulted in a slightly later crossing-over (at about between the TFS and ENV compared to the Hilbert envelope extraction in both the sentence-sentence and music-music chimera trials. As mentioned in A, the Hilbert envelope extraction results in better representation of the fine changes in the signal envelope compared to the rectified low-pass filter approach. This may explain why the subjects, on average, identified hearing the envelope more clearly at lower bands in the Hilbert curves than they did in the 16 and 160 Hz curves.
- d) The main finding that is consistent between the Smith study and ours, despite our *much* less extensive testing, is the general observation that the envelope of the waveform appears to be more useful in identifying words/speech patterns in sentences while the fine structure of the waveform seems to be more useful in identifying music. However, at very low numbers of bands, it appears fine structure is more important for sentence identification. In music, fine structure seems to be most useful at most frequency bands, until the transition at ~25-30 bands, where the envelope starts to become more useful.
- e) Modern cochlear implants use a similar filter-bank paired with envelope extraction approach to the methods we used in this project. Current implants transcode information in speech quite well, and give users the ability to successfully listen and communicate. However, these implants do not transcode the temporal fine structure of sound because of the low-pass filtering or Hilbert transform used to derive the envelope for each band. This is unfortunate, because in music, the temporal fine structure is important in conveying *timbre*. Looking at the frequency spectrum of different instruments playing the same note shows small peaks beyond the fundamental frequency of that note. These small variations that exist in the TFS are what give instruments their unique sounds, and are unable to be properly replicated in CIs.

Improving the ability of CI users to hear music or other sounds that have useful information in the TFS is a good next step in cochlear implant development. Most prospects for improving TFS delivery to CIs involve better replication of the TFS. One particular example is found in a 2012 study by Li et al. (Rubinstein Lab) that used a Harmonic Single Sideband Encoder (HSSE) to attempt to better convey TFS. The basic summary of this method is to identify and relay the harmonic content of a given signal to the output channels, which results in more TFS representation of a signal than methods such as continuous-interleaved-sampling (CIS).

### 3 Part C | Consiering a technical artifact with chimera synthesis and its potential implications