Digital Signal Processing Lab
Experiment 6(a)
Chimaeric sounds reveal dichotomies in auditory perception

Name: Anand Jhunjhunwala
Roll No: 17EC30041
Group No: 63

Objective:

The goal of the assignment is to construct a summary of the paper titled "Chimaeric sounds reveal dichotomies in auditory perception" and try to gain an understanding of aspects of sound features, particularly the location of the sound source and identity (recognition) of the sound that are derived from the envelope (slow variations in the sound) and the fine structure (higher frequency fluctuations) in the sound.

Summary:

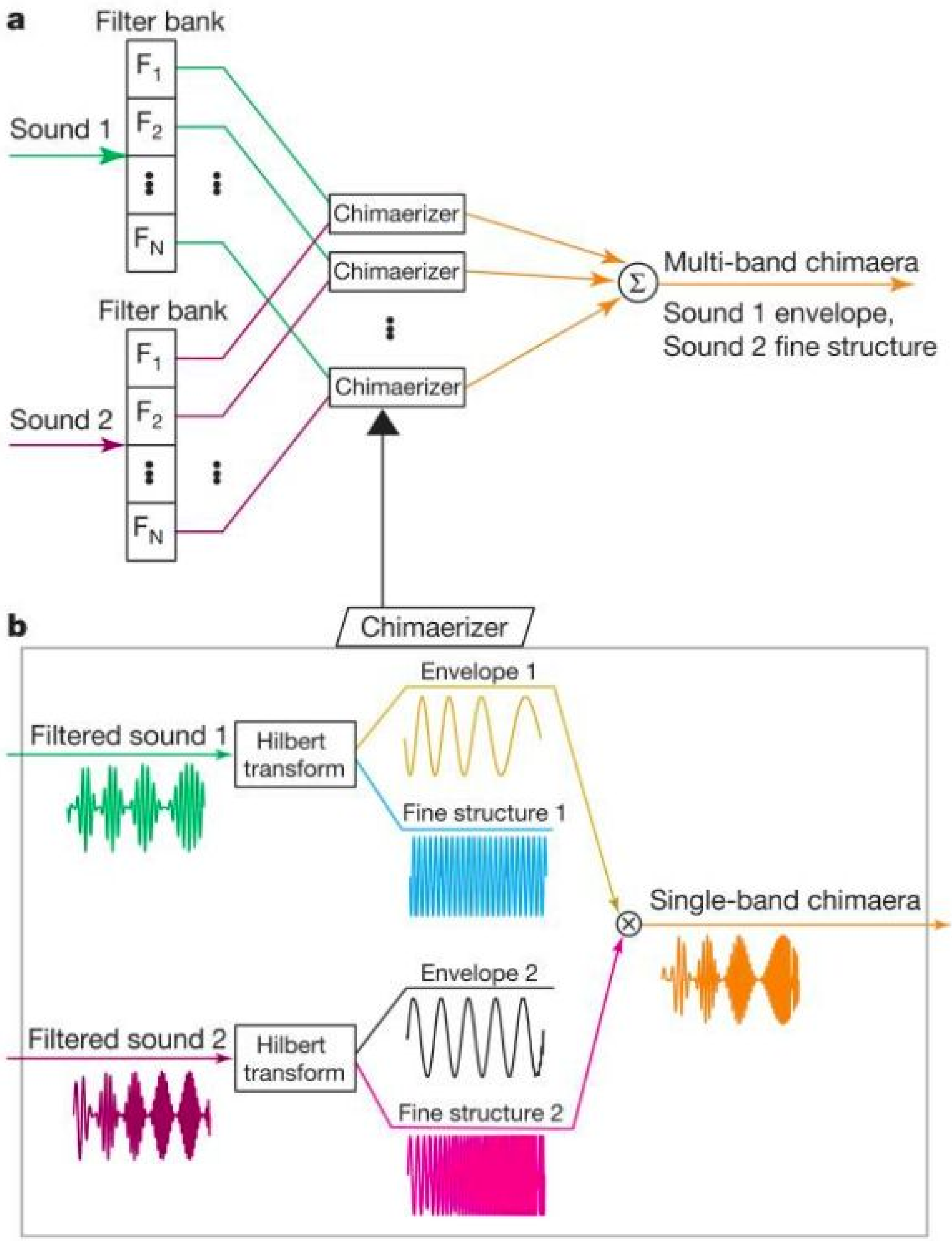
In the paper, the authors state that neurons in the auditory brainstem sensitive to features like sound envelop (slow variations in the sound) and the fine structure (higher frequency fluctuations) in the sound apart from its frequency components. So they investigate the relative perceptual importance of envelope and fine structure, by synthesizing stimuli that called '**auditory chimaeras**', which have the envelope of one sound and the fine structure of another.

The main theme of the paper is to investigate the relative importance of fine structure and envelop in determining sound features mainly, location, and preception.

<u>Synthezation of auditory chimaeras.</u>

To synthesize auditory chimaeras, two sound waveforms are used as inputs.
Step1: A bank of band-pass filters is used to split each sound into 1 to 64 complementary frequency bands spanning the range 80–8,820 Hz.
Step2: The output of each filter is factored into its envelope and fine structure using the Hilbert transform.
Step3: The envelope of each filter output from the first sound is then multiplied by the fine structure of the corresponding filter output from the second sound.
Step4: These products are finally summed over all frequency bands to produce an auditory chimaera that is made up of the envelope of the first sound and the fine structure of the second sound in each band. (see block diagram below)

Note: The primary variable in this study is the number of frequency bands, which is inversely related to the width of each band.



Simulation Study 1:

Part a:

In this study, the author tries to examine the effect of the envelope and fine structures in speech perception. For the study(a), they used speech-noise chimaeras. In which envelope of a speech sentence is combined with the fine structure of noise and vise versa.

Experiment condition
Subjects listened to the processed sentences and were instructed to type the words they heard into a computer. Each subject listened to a total of 273 speech chimaeras with an additional seven for training. Speech reception was measured as percentage words correct. 'The', 'a' and 'an' were not scored. When speech–speech chimaeras were used, each word in a subject's response could count for either a sentence in the envelope or in the fine structure.

Results:

1) Speech reception is highly dependent on the number of frequency bands used for the synthesis.
2) When speech information is contained solely in the envelope, speech reception is poor with one or two frequency bands and improves as the number of bands increases. Good performance (>85% word recognition) is achieved with as few as four frequency bands.
3) When speech information is only contained in the fine structure, speech reception is generally better with fewer frequency bands. The best performance is achieved with two bands and performance then deteriorates as the number of bands increases until, with eight or more bands, there is essentially no speech reception.

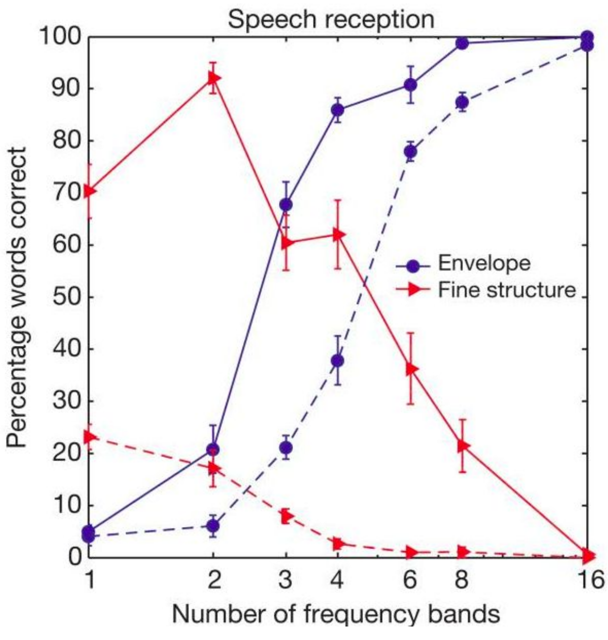
NOTE: The facts are also verified in our experiment 4.

Part b:
In this study, the author did the same investigation but with speech-speech chimaeras constructed with two distinct utterances.

Results:

1) Even though speech–speech chimaeras are constructed with two distinct utterances, listeners almost invariably heard words from only one of the two sentences.
2) Speech reception based on the fine structure was much poorer with speech–speech chimaeras than with speech–noise chimaeras.
3) Speech reception based on envelope information was also degraded, but not as severely, and performance still exceeded 80% with eight or more frequency bands.

Result Graph:



Note: In the above graph, the solid line is for speech-noise chimaeras and the dotted line is for speech-speech chimaeras.

Conclusion:
Speech perception is highly dominated by envelope then fine structure for the higher band and vice-versa for 1 or 2 bands.

Simulation Study 2:

In this study, the author tries to examine the effect of the envelope and fine structures in melody recognition. To examine this the author used chimaeras based on two different melodies, one in the envelope and the other in the fine structure.

This study is performed to analyze the pitch perception of complex harmonic sounds.

Experiment condition
Each subject selected ten melodies that he/she was familiar with. Each melody was taken from a set of 34 simple melodies with all rhythmic information removed, consisting of 16 equal-duration notes and synthesized with MIDI software that used samples of a grand piano. During the experiment, subjects selected from their own list of ten melodies which one(s) they heard on each trial. Melodies were scored as percentage correct even when subjects reported multiple melodies in a single trial without penalty for incorrect responses.

Results:

1) Melody-melody chimaeras show a reversal in the relative importance of envelope and fine structure when compared to speech-speech chimaeras.

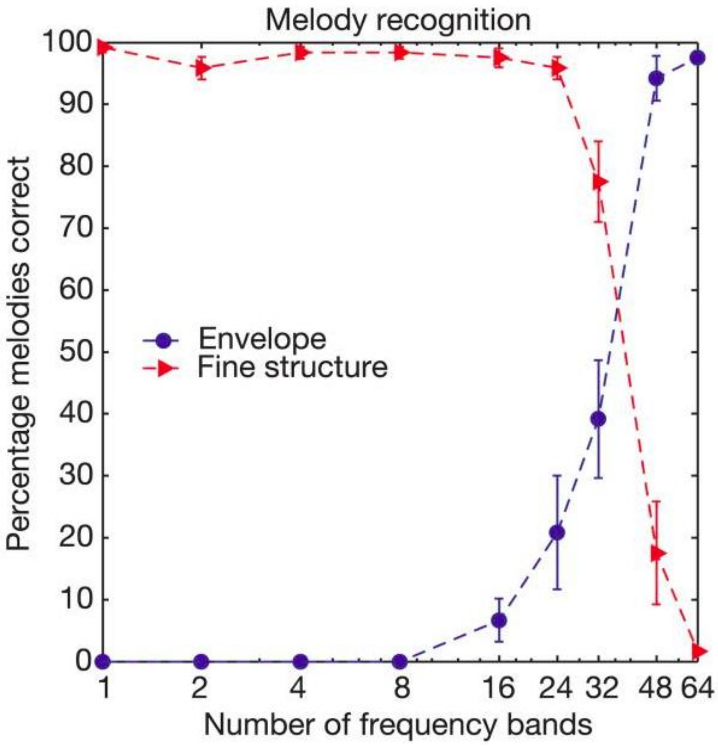
2) Listeners always heard the melody based on the fine structure with up to 32 frequency bands.

3) With 48 and 64 frequency bands, however, they identified the envelope-based melody more often than they did the melody based on fine structure.

4) Subjects sometimes reported hearing two melodies and often picked the two melodies represented in the envelope and the fine structure, respectively.

5) The crossover point, where the envelope begins to dominate over the fine structure, occurs for a much higher number of frequency bands (about 40) for melody–melody chimaeras than it does for speech–speech and speech–noise chimaeras.

Result graph:



Studies on Sound Localization:

Sound localization in the horizontal plane is based on interaural differences in time and level. Interaural time differences (ITD) are the dominant cue for low-frequency sounds such as speech.

Construction of dichotic chimaeras.

Step 1: A delay of 700 μs was introduced into either the right or left channel of each sentence to create ITDs that would produce completely lateralized sound images.

Step 2: Sentence with an ITD pointing to the right was combined with a sentence having an ITD pointing to the left to produce a chimaera with its envelope information pointing to one side and its fine structure pointing to the other side.

<u>Simulation Study 3:</u>

In this study two types of dichotic chimaeras were constructed, one using the same sentence for both the envelope and fine structure, and the other using different sentences. These chimaeras were used to study the lateralization of sound.

Experiment condition

Subjects used a seven-point scale to rate the lateral position of the sound image inside the head. This scale ranges from -3 to +3, with -3 corresponding to the left ear and +3 to the right ear. Lateralization scores were averaged for each condition. In addition, subjects had to select which of two possible sentences they heard, one choice corresponding to the envelope and the other one to the fine structure.

Results:

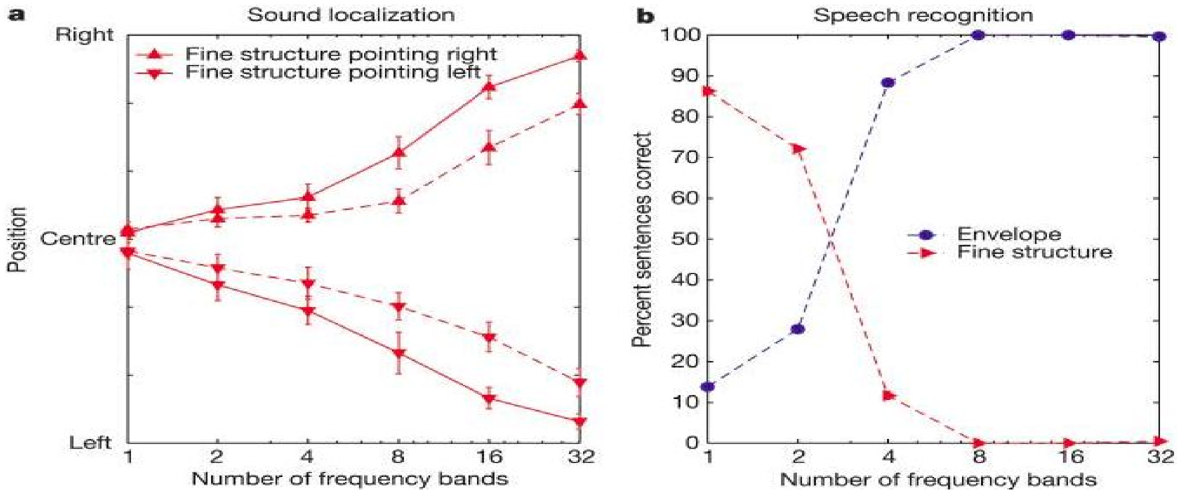
1) The lateralization of dichotic chimaeras was always based on the ITD of the fine structure.
2) Lateralization improved with an increasing number of bands.
3) Dichotic chimaeras based on the same sentence in the envelope and fine structure were more easily lateralized than those based on two different sentences.

Note: When dichotic chimeras based on different sentences were presented, listeners were asked to pick which of the two sentences they heard in addition to reporting the lateral position of the sound image.

Result:

1) Consistent with our results for speech–speech chimaeras, subjects most often heard the sentence based on the fine structure with one and two frequency bands.
2) The subject heard the sentence based on the envelope for four or more bands. With eight or more frequency bands, subjects clearly identified the sentence based on the envelope but lateralized the speech to the side to which the fine structure was pointing.

Result Graph



Conclusion:

The fine structure determines ‘where’ the sound is heard, whereas the envelope determines ‘what’ sentence is heard. In this respect, auditory chimaeras are consistent with evidence for separate ‘where’ and ‘what’ pathways in the auditory cortex.

<u>About the subjects participated in the study.</u>

1) Six native speakers of American English with normal hearing thresholds participated in each part of the study. I.e (Speech reception, melody recognition, and lateralization), each test is conducted separately.
2) Within each individual experiment, the order of all conditions was randomized. Stimuli were presented in a soundproof booth through headphones at a root-mean-square sound pressure level of 67 dB.

Mathematics of Stimulus synthesis

The perfect-reconstruction digital filter banks used for chimaera synthesis spanned the range 80–8,820 Hz, spaced in equal steps along the cochlear frequency map (nearly logarithmic frequency spacing). For example, with six bands, the cutoff frequencies were 80, 260, 600, 1,240, 2,420, 4,650 and 8,820 Hz. The transition over which adjacent filters overlap significantly was 25% of the bandwidth of the narrowest filter in the bank (the lowest in frequency). Thus, for the six band case, each filter transition was 45 Hz wide.

To compute the envelope and fine structure in each band, the analytic signal was used $s(t) = s_r(t) + is_i(t)$, where $s_r(t)$ is the filter output in one band, $s_i(t)$ the Hilbert transform of $s_r(t)$, and $i=\sqrt{-1}$. The Hilbert envelope is the magnitude of the analytic signal, $a(t)=\sqrt{(s_r^2(t)+s_i^2(t))}$. The fine structure is $\cos \phi(t)$, where $\phi(t) = \arctan(s_i(t)/s_r(t))$ is the phase of the analytic signal. The original signal can be reconstructed as $s_r(t) = a(t) \cos \phi(t)$. In practice, the Hilbert transform was combined with the band-pass filtering operation using complex filters whose real and imaginary (subscripts r, i) parts are in quadrature

Conclusion about Cochlear implants:

Cochlear implants are prosthetic devices that seek to restore hearing in the profoundly deaf by stimulating the auditory nerve via electrodes inserted into the cochlea.

Current technology in cochlear implants discard the fine time structure and present only about six to eight bands of envelope information. As we have seen in the above results that the fine structure of sound plays an important role in pitch perception and sound localization.

So, the inclusion of fine structure in these implants will help the patient to improve its pitch perception and sensitivity to ITD.

1) With better pitch perception the patient will be able to appreciate music, as well as will be able to understand tonal languages, such as Mandarin Chinese, where the pitch is used to distinguish different words.
2) Better ITD sensitivity may help the increasing number of patients with bilateral cochlear implants in taking advantage of binaural cues that normal-hearing listeners use to distinguish speech among competing sound sources.