NCSI – Project 1

-Aditya Sinha (14EC10002)

PART A

1. Responses to tones (rate representation)

In **Fig. 1-2**, we observe that the peak of the tuning curves lies near the BF for both the ANFs. Furthermore, increasing sound levels increases the peak and the sharpness of the tuning curve.

Upon plotting the rate vs. intensity for the ANFs in **Fig. 3**, we observe that the peak rate averaged over 20 repetitions increases with increasing sound levels as a monotonic function. In this, above spontaneous rate, the rate of firing of ANF with BF=500Hz is always greater than that of BF=4kHz ANF. This difference further increases as intensity increases.

2. Responses to speech (rate representation)

First, we separate out the steady state *"**a**"* in *"five women play b**a**sketball"*

Next, we calibrate the sound levels using this steady state speech section and we note the firing rates at different intensity levels. For ANF with BF=500Hz, the rate vs. intensity plot for this vowel is shown in **Fig. 4**.

Based on this, we choose the three sound levels (20,50,80 dB SPL) such that one lies just above threshold, one lies in dynamic range and the third lies near saturation regime.

At these 3 sound levels, we send the entire speech signal 50 times and calculate the spike trains, thus getting a psth of firing with bin size 0.01ms(1/Fs).

Next, we plot the spectrogram of the speech signal. We compare this with our own spectrogram kind of graph, with firing rate for different BF ANFs is used as a proxy for the spectrogram. This process is repeated for all the three intensities and the results are given in **Fig. 5-10**.

We observe that barring the logarithmic nature of the frequency axis in the rate proxy spectrogram and the color rescaling due to lesser frequency range, we obtain a plot similar to the spectrogram. At medium window size, we see the same gaps in the plot that signify pauses between words and syllables. However, this similarity is maximum in the 50dB SPL case and becomes lesser and lesser as we go out of the dynamic range of the ANF. Across sound levels, we can see the warmth and contrast of spectrogram increasing with increasing intensities.

3. Responses to speech (fine timescale representation)

For this part, we group the earlier obtained psth into fine timescale bins of 0.1 ms each (binlen=binsize*Fs=10 data pts).

We then take a moving window STFT with window size 12.8 ms and 50% overlap. Taking argmax of each of the fft peaks, we get a measure of the dominant frequency in phase locking for that ANF in that particular time window. We then repeat this for different time slices, different BF ANFs and overall different intensities to get phase locking plots (as shown by the asterisk plots in **Fig. 11-13**).

We observe that significant phase locking only appears to happen for a certain range of BF ANFs, and that too only at the intensity in the dynamic range. At higher frequencies, ANF phase locks to low f envelope.

At 50dB SPL, ANFs with BF above 2.8kHz or below 1kHz do not show sufficient phase locking over all times.

Furthermore, there is a sharp variation from mean dominant frequency in the gap regions of the speech signal, indicating a sudden change in intensity and psth of firing.

PART B

Upon presenting the reconstructed speech signals to an unbiased listener, it was observed that there was some sense of intelligibility by 4 bands, and by 8 bands, the listener could easily comprehend the words in the sentence.

Next, using 1 band and 4 band reconstructed speech signals using temporal cues, we repeated parts A2 and A3 to gain insight into the nature of the reconstructed signal.

Similar to A2 and A3, the vowel “ah” was extracted from the reconstructed speech signal. The rate vs. intensity plot for both 1 band and 4 bands is shown in **Fig. 14-15**. We observe that the rate of firing varies a little more erratically with intensity for 1 band whereas it follows a similar trend for 4 bands, but ultimately, the monotonicity is maintained in both cases. The erratic variation is due to the loss of spectral information in the speech signal.

Rescaling the sound level of the entire reconstructed signal and then passing it to the neuron, spike trains were obtained, which were then used to construct rate proxy spectrograms, to compare with the actual spectrogram of the reconstructed signal (**Fig. 16-27**). From the spectrograms, along the y-axis, we observe warm regions around band centers, giving 1 and 4 regions for 1 and 4 bands respectively. These regions become a little fuzzy and spread out with intensity, but nevertheless, the distinction is maintained. However, within each region, there isn’t a lot of variation to be seen, indicating removal of spectral information. Along the x-axis, we see the effect of preservation of temporal cues, with the earlier seen gaps still visible in the spectrogram. The similar behaviour of the rate proxy spectrogram shows that the neuron is also able to understand the reconstructed speech to an extent. This is seen best in 4 bands, 50dB SPL, 4ms window.

Finally, we do the ANF phase locking analysis for the reconstructed signal (**Fig. 28-33**). We observe best phase locking in 4 bands, 50dB SPL case.

Having made this comparison between 1 and 4 bands via simulation and analysing various parameters, we observe that the 4 bands reconstructed signal tests significantly better than the 1 band one. This is concurrent with the observation of the listener, further strengthening the hypothesis presented by the paper: *Shannon et al* – “*Speech Recognition with Primarily Temporal Cues”*

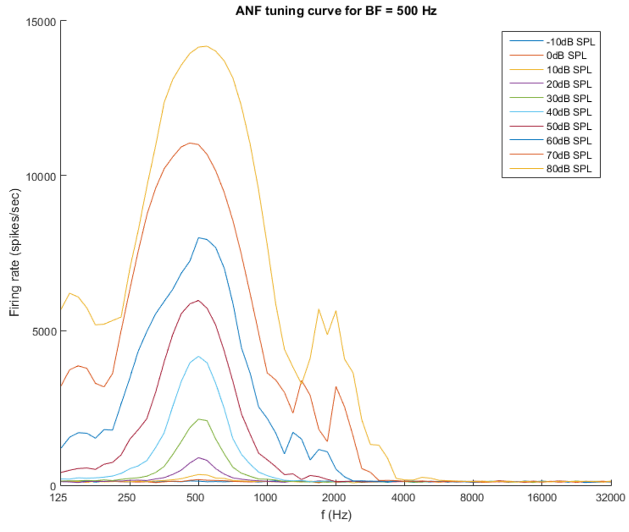


Fig. 1

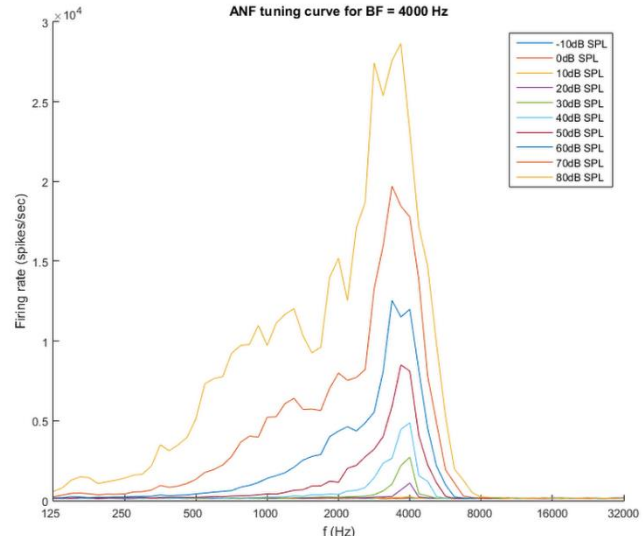


Fig. 2

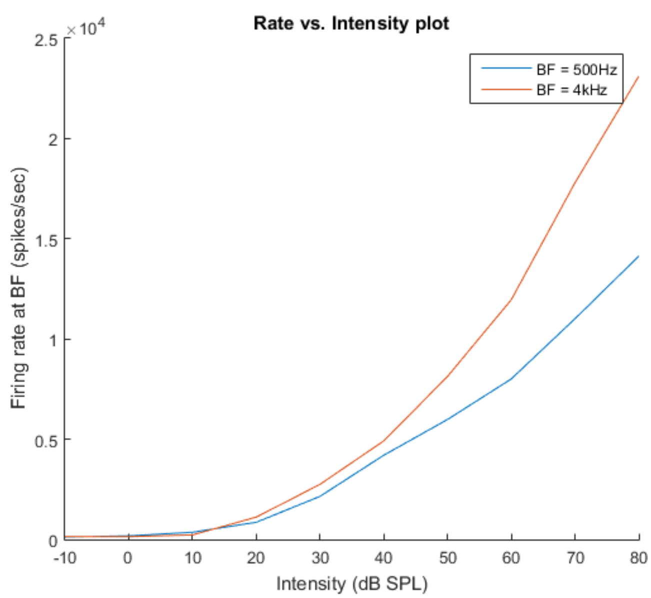


Fig. 3

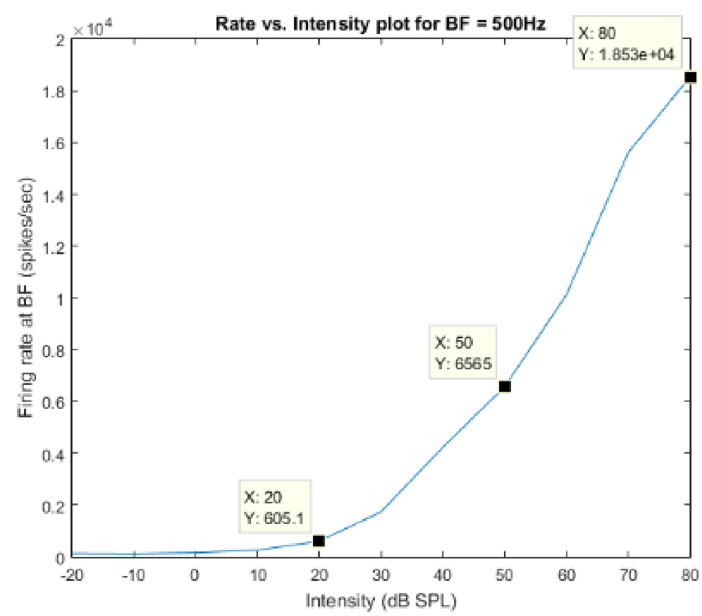
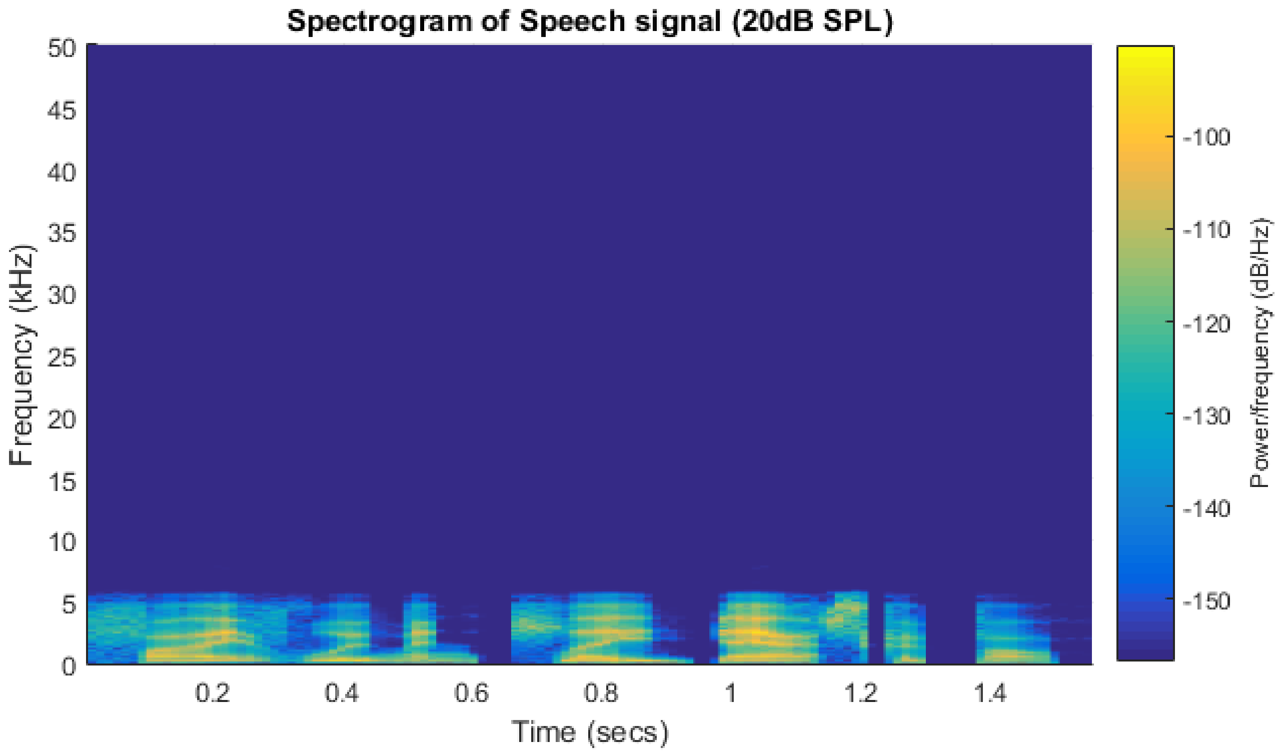
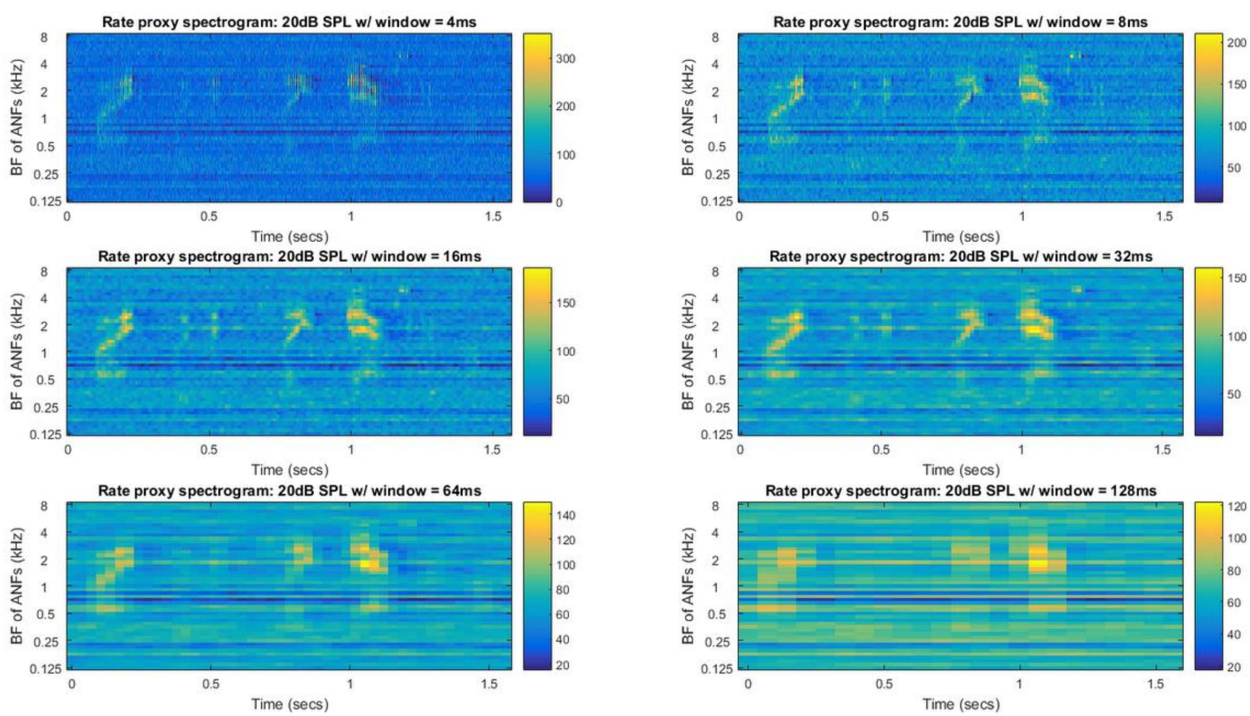


Fig. 4



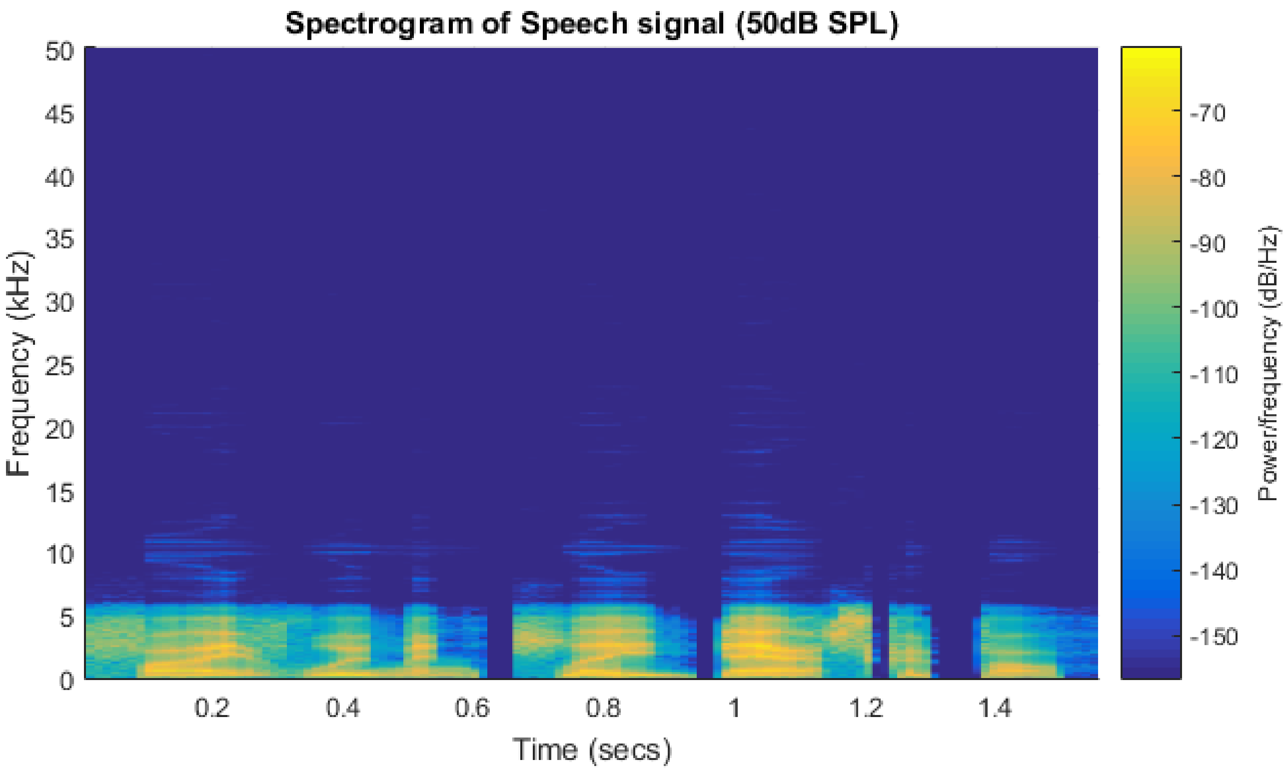
Spectrogram of Speech signal (20dB SPL)
Frequency (kHz)
Time (secs)
Power/frequency (dB/Hz)

Fig. 5



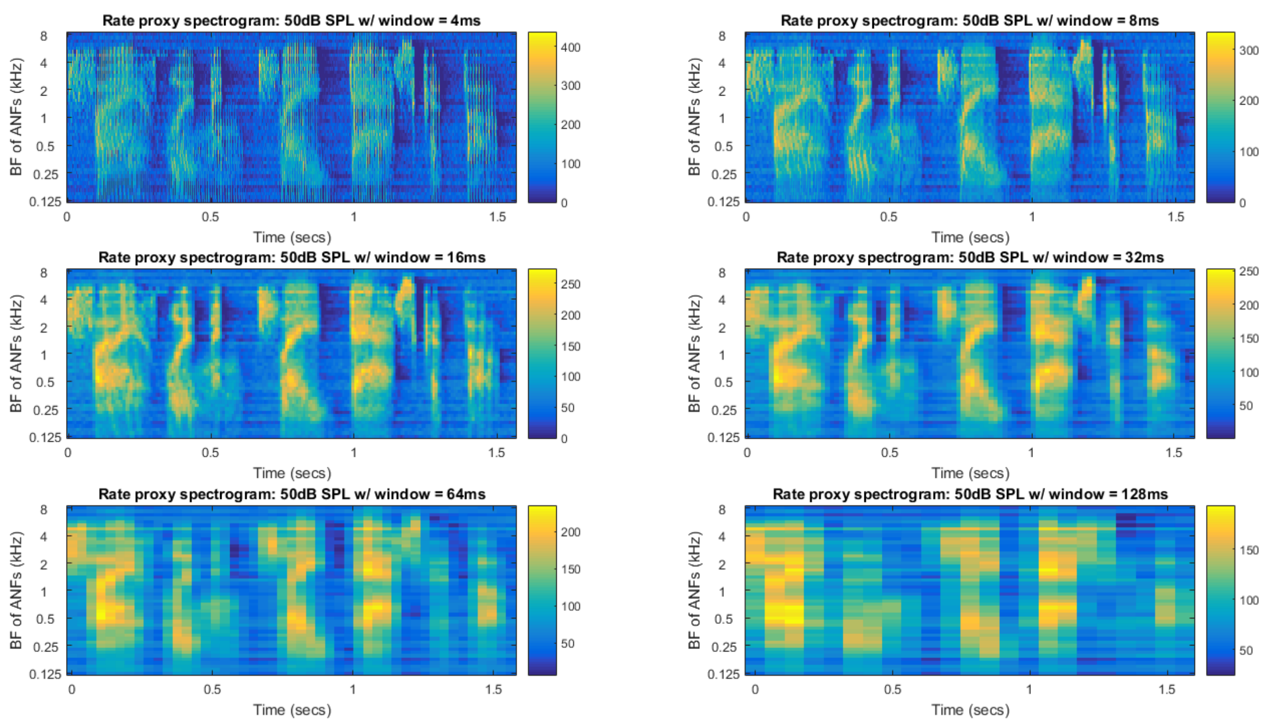
Rate proxy spectrogram: 20dB SPL w/ window = 4ms
Rate proxy spectrogram: 20dB SPL w/ window = 8ms
Rate proxy spectrogram: 20dB SPL w/ window = 16ms
Rate proxy spectrogram: 20dB SPL w/ window = 32ms
Rate proxy spectrogram: 20dB SPL w/ window = 64ms
Rate proxy spectrogram: 20dB SPL w/ window = 128ms
BF of ANFs (kHz)
Time (secs)

Fig. 6



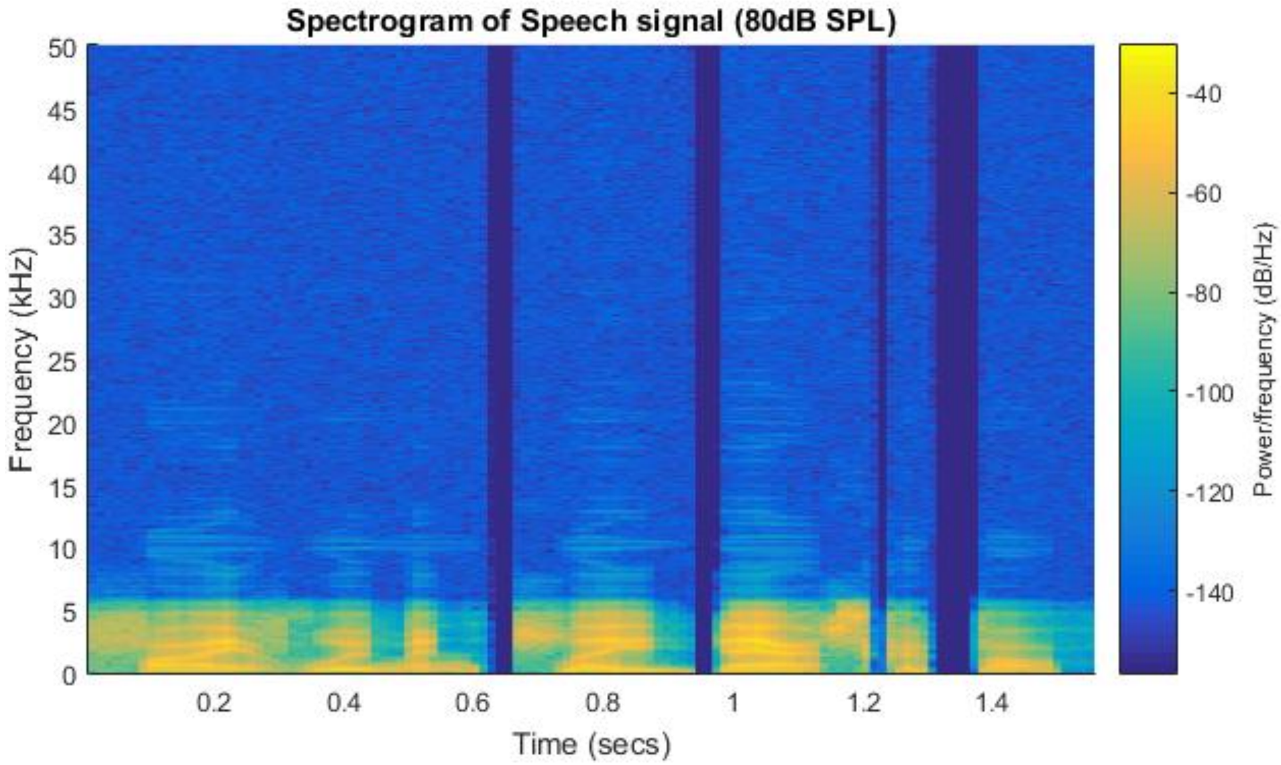
Spectrogram of Speech signal (50dB SPL)
Frequency (kHz)
Time (secs)
Power/frequency (dB/Hz)

Fig. 7



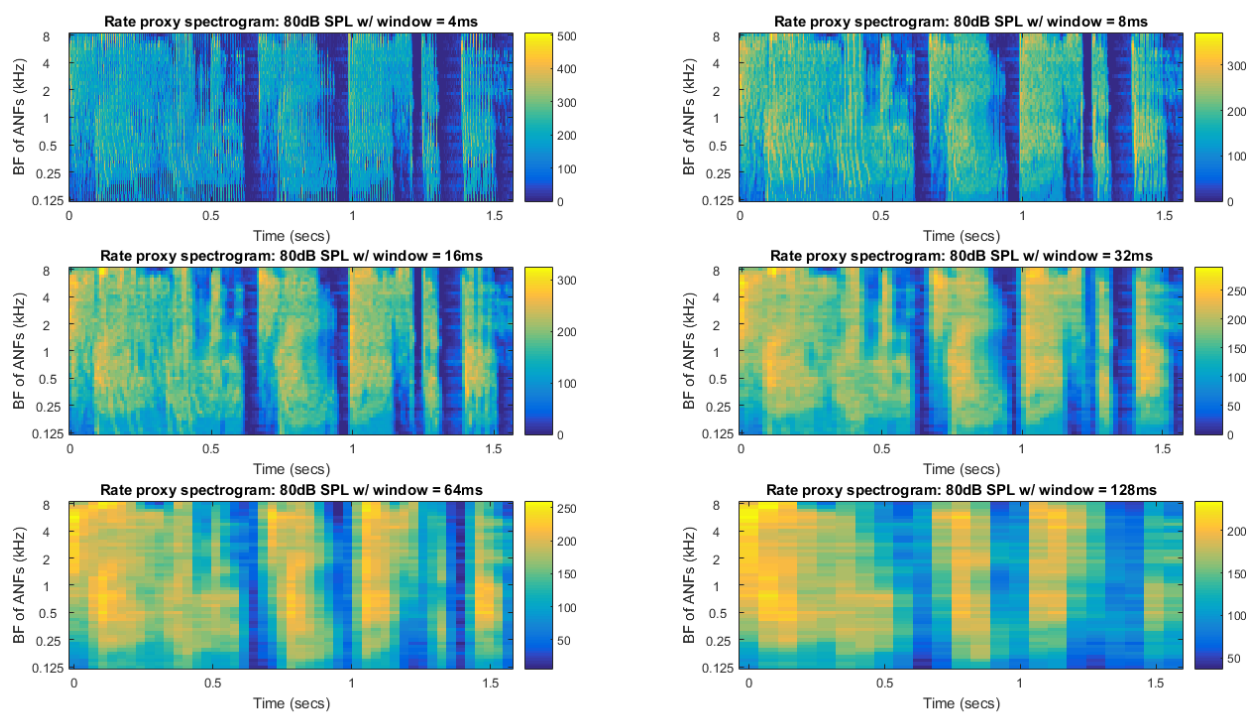
Rate proxy spectrogram: 50dB SPL w/ window = 4ms
Rate proxy spectrogram: 50dB SPL w/ window = 8ms
Rate proxy spectrogram: 50dB SPL w/ window = 16ms
Rate proxy spectrogram: 50dB SPL w/ window = 32ms
Rate proxy spectrogram: 50dB SPL w/ window = 64ms
Rate proxy spectrogram: 50dB SPL w/ window = 128ms
BF of ANFs (kHz)
Time (secs)

Fig. 8



Spectrogram of Speech signal (80dB SPL)
Frequency (kHz)
Time (secs)
Power/frequency (dB/Hz)

Fig. 9



Rate proxy spectrogram: 80dB SPL w/ window = 4ms
Rate proxy spectrogram: 80dB SPL w/ window = 8ms
Rate proxy spectrogram: 80dB SPL w/ window = 16ms
Rate proxy spectrogram: 80dB SPL w/ window = 32ms
Rate proxy spectrogram: 80dB SPL w/ window = 64ms
Rate proxy spectrogram: 80dB SPL w/ window = 128ms
BF of ANFs (kHz)
Time (secs)

Fig. 10

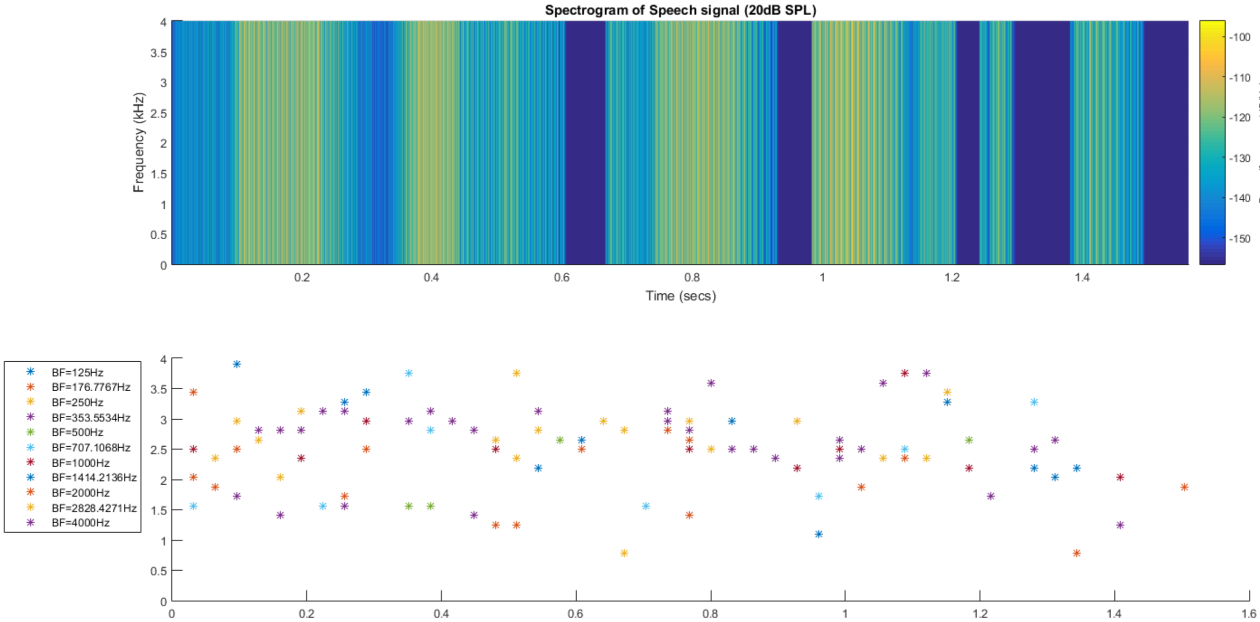

Spectrogram of Speech signal (20dB SPL)
Frequency (kHz)
Time (secs)
Power/frequency (dB/Hz)
BF=125Hz
BF=176.7767Hz
BF=250Hz
BF=353.5534Hz
BF=500Hz
BF=707.1068Hz
BF=1000Hz
BF=1414.2136Hz
BF=2000Hz
BF=2828.4271Hz
BF=4000Hz

Fig. 11

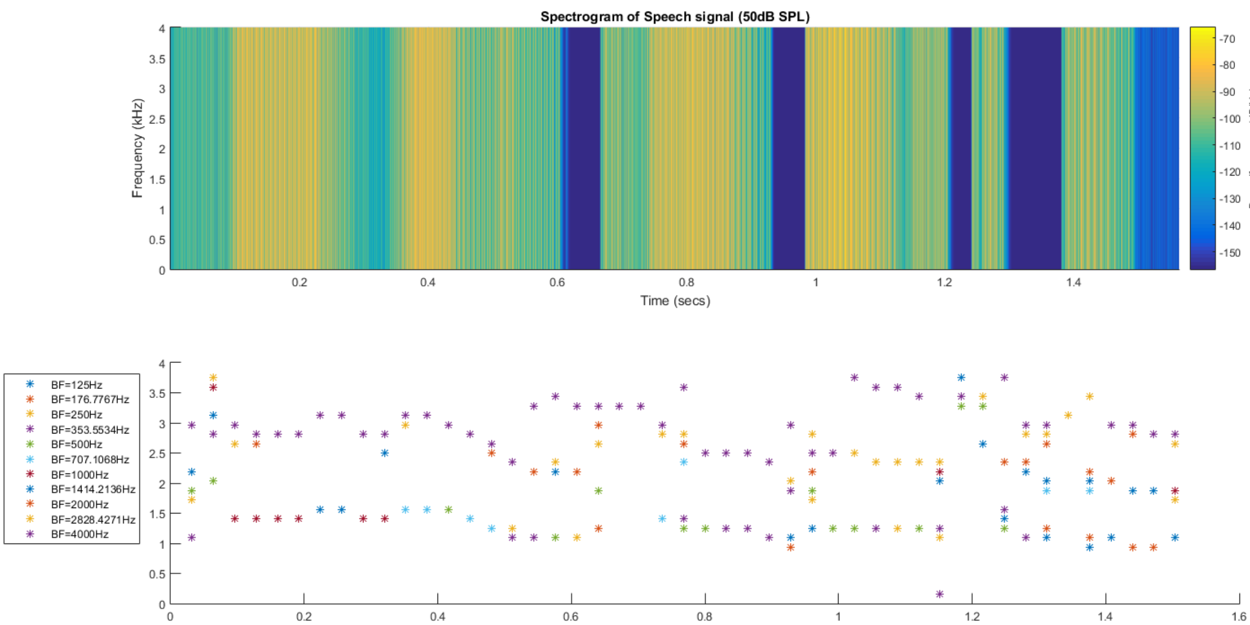

Spectrogram of Speech signal (50dB SPL)
Frequency (kHz)
Time (secs)
Power/frequency (dB/Hz)
BF=125Hz
BF=176.7767Hz
BF=250Hz
BF=353.5534Hz
BF=500Hz
BF=707.1068Hz
BF=1000Hz
BF=1414.2136Hz
BF=2000Hz
BF=2828.4271Hz
BF=4000Hz

Fig. 12

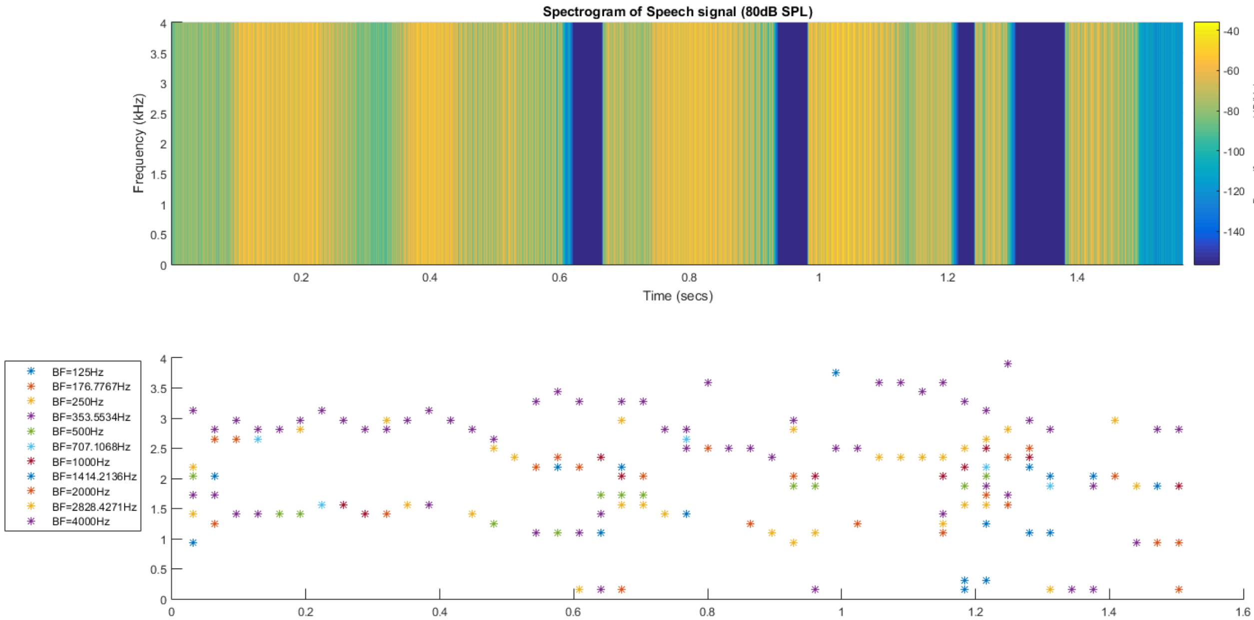


Fig. 13

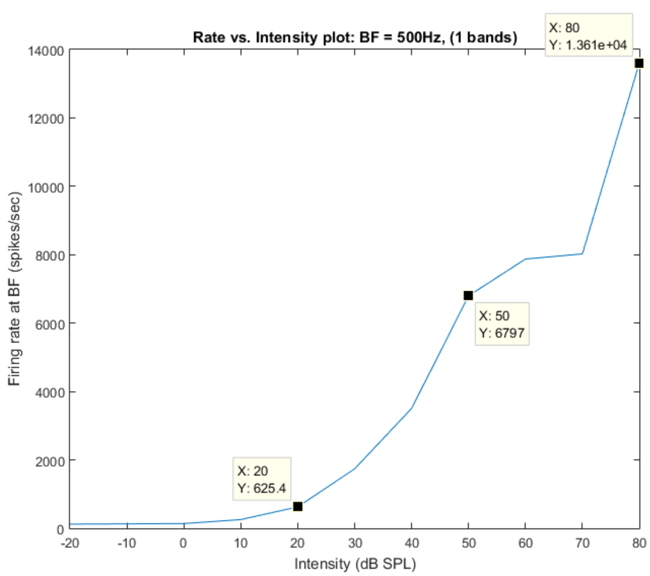


Fig. 14

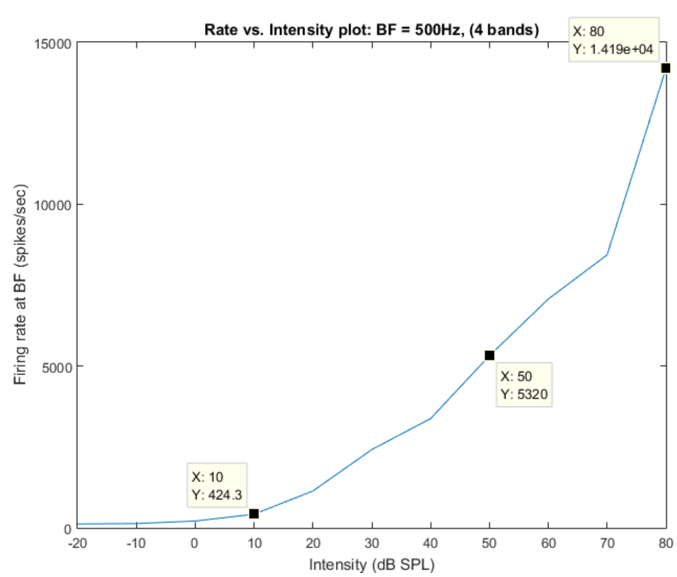
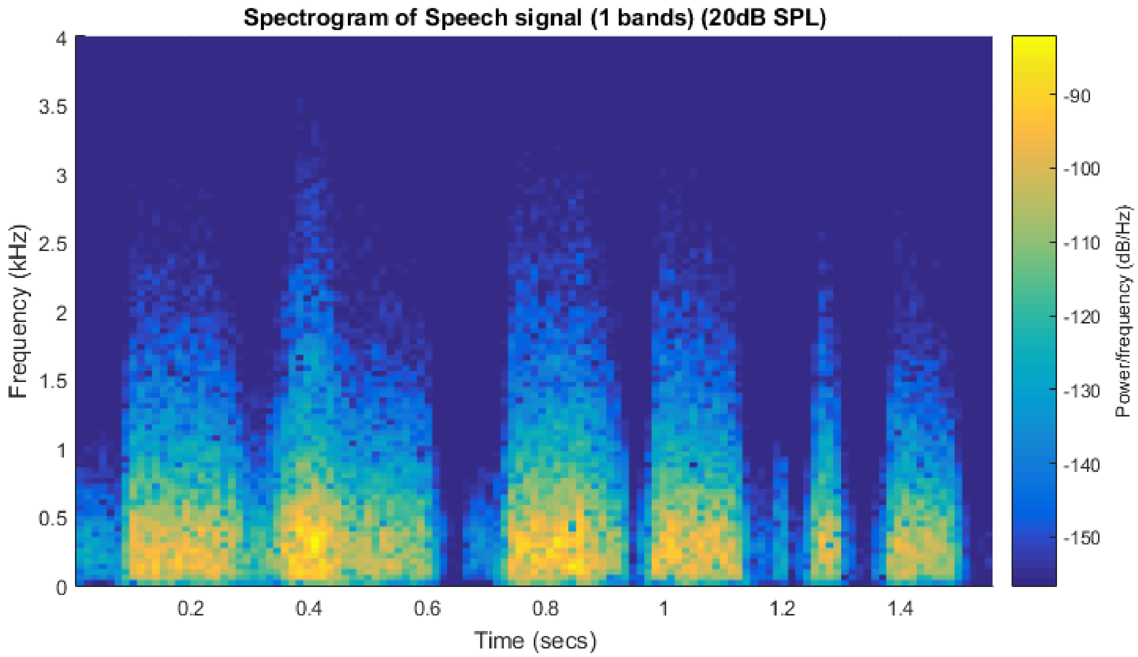
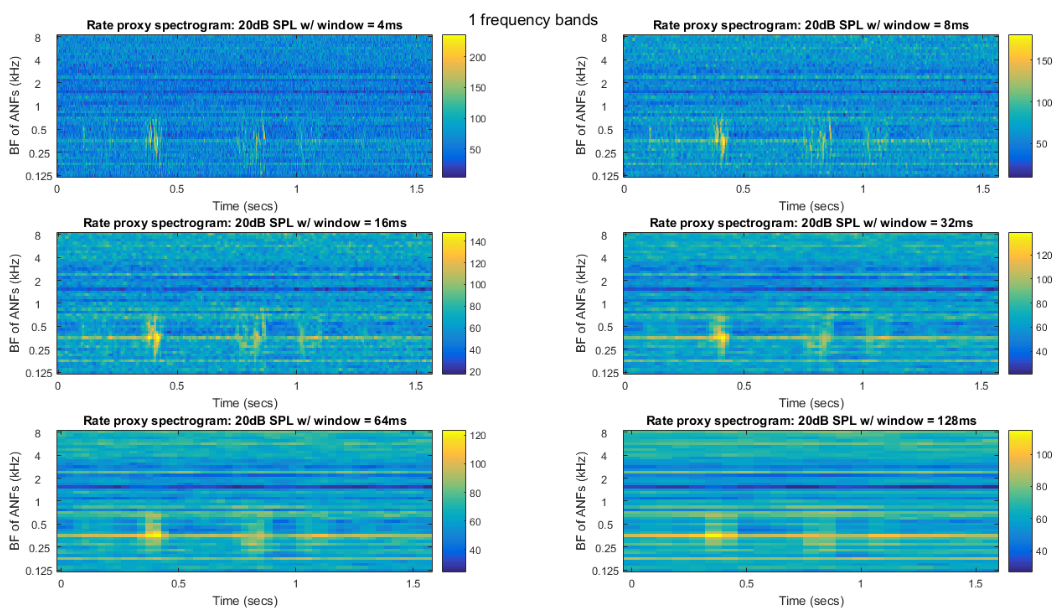


Fig. 15



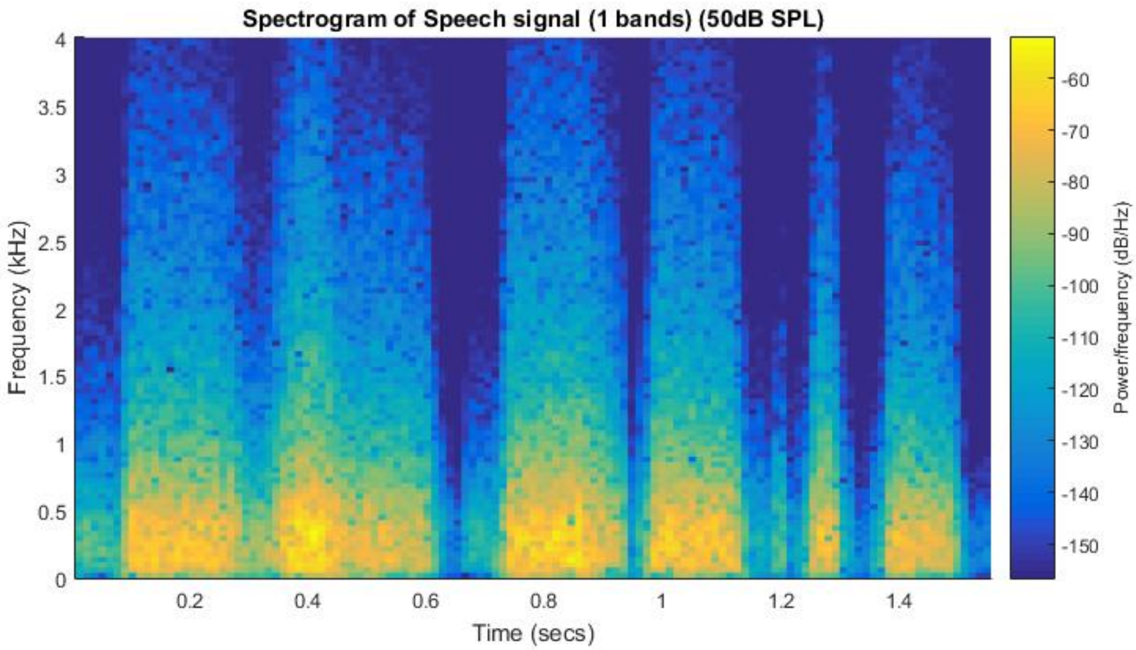
Spectrogram of Speech signal (1 bands) (20dB SPL)
Frequency (kHz)
Time (secs)
Power/frequency (dB/Hz)

Fig. 16



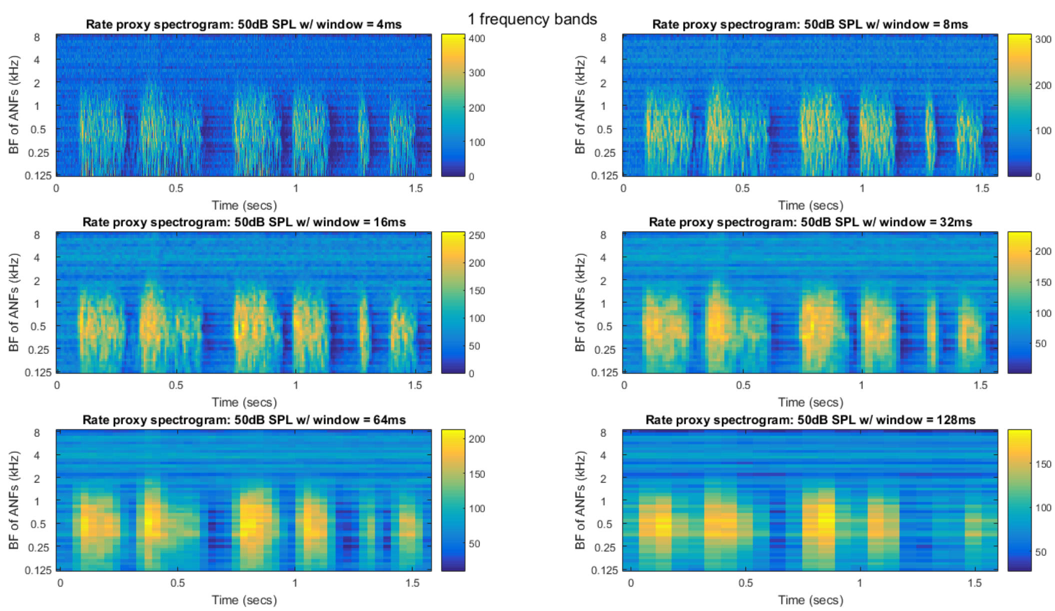
1 frequency bands
Rate proxy spectrogram: 20dB SPL w/ window = 4ms
Rate proxy spectrogram: 20dB SPL w/ window = 8ms
Rate proxy spectrogram: 20dB SPL w/ window = 16ms
Rate proxy spectrogram: 20dB SPL w/ window = 32ms
Rate proxy spectrogram: 20dB SPL w/ window = 64ms
Rate proxy spectrogram: 20dB SPL w/ window = 128ms
BF of ANFs (kHz)
Time (secs)

Fig. 17



Spectrogram of Speech signal (1 bands) (50dB SPL)
Frequency (kHz)
Time (secs)
Power/frequency (dB/Hz)

Fig. 18



1 frequency bands
Rate proxy spectrogram: 50dB SPL w/ window = 4ms
Rate proxy spectrogram: 50dB SPL w/ window = 8ms
Rate proxy spectrogram: 50dB SPL w/ window = 16ms
Rate proxy spectrogram: 50dB SPL w/ window = 32ms
Rate proxy spectrogram: 50dB SPL w/ window = 64ms
Rate proxy spectrogram: 50dB SPL w/ window = 128ms
BF of ANFs (kHz)
Time (secs)

Fig. 19

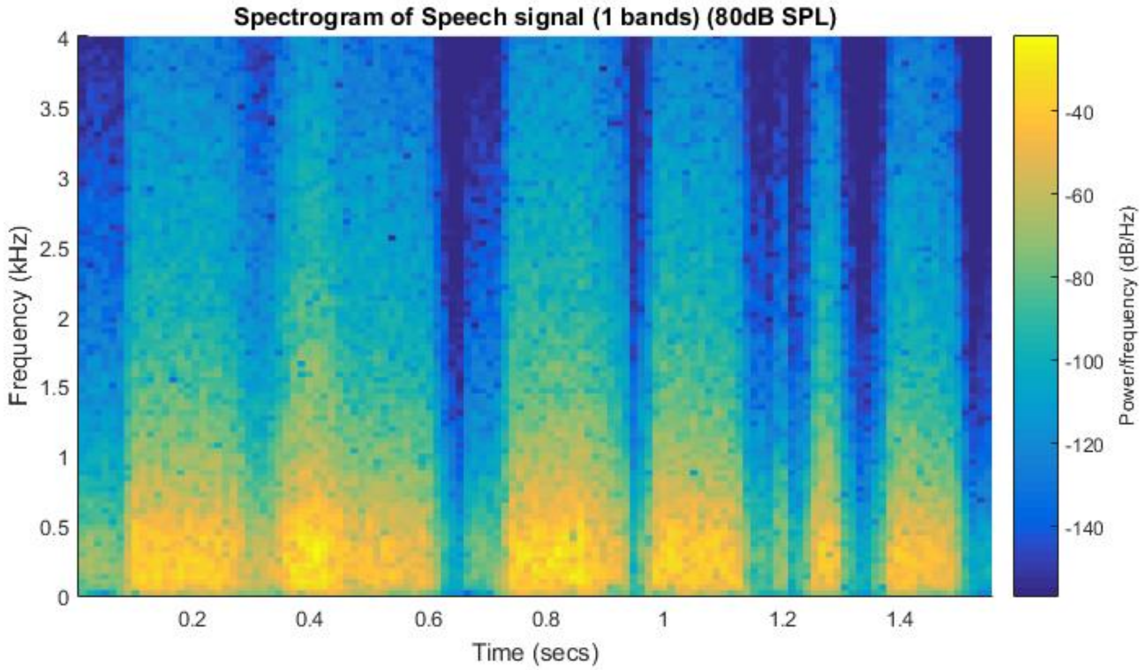

Spectrogram of Speech signal (1 bands) (80dB SPL)
Frequency (kHz)
4
3.5
3
2.5
2
1.5
1
0.5
0
0.2
0.4
0.6
0.8
1
1.2
1.4
Time (secs)
-40
-60
-80
-100
-120
-140
Power/frequency (dB/Hz)

Fig. 20

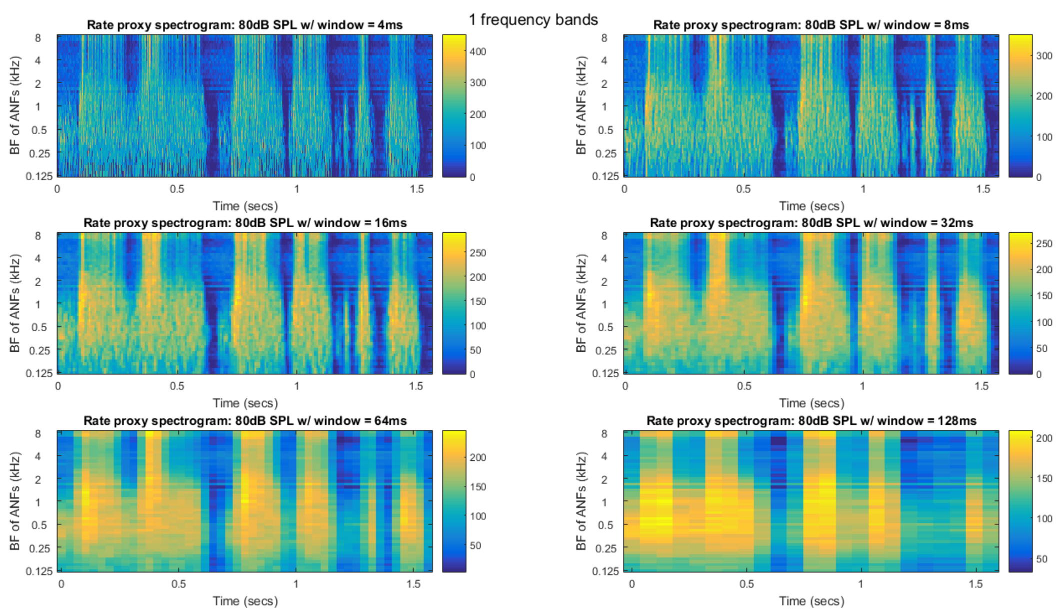
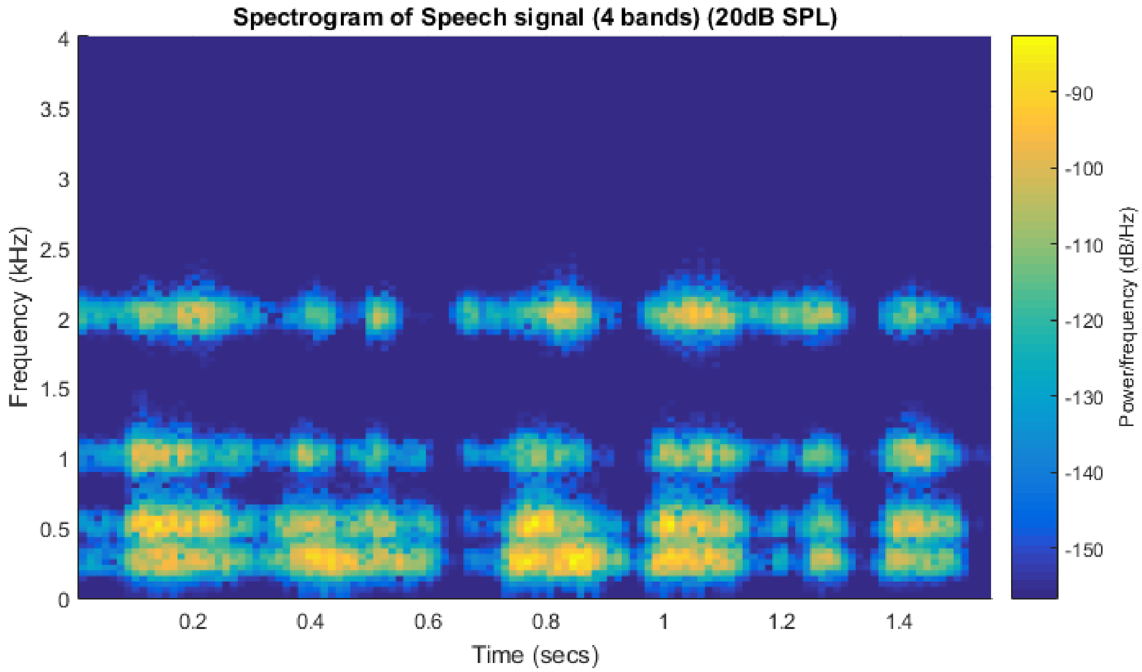
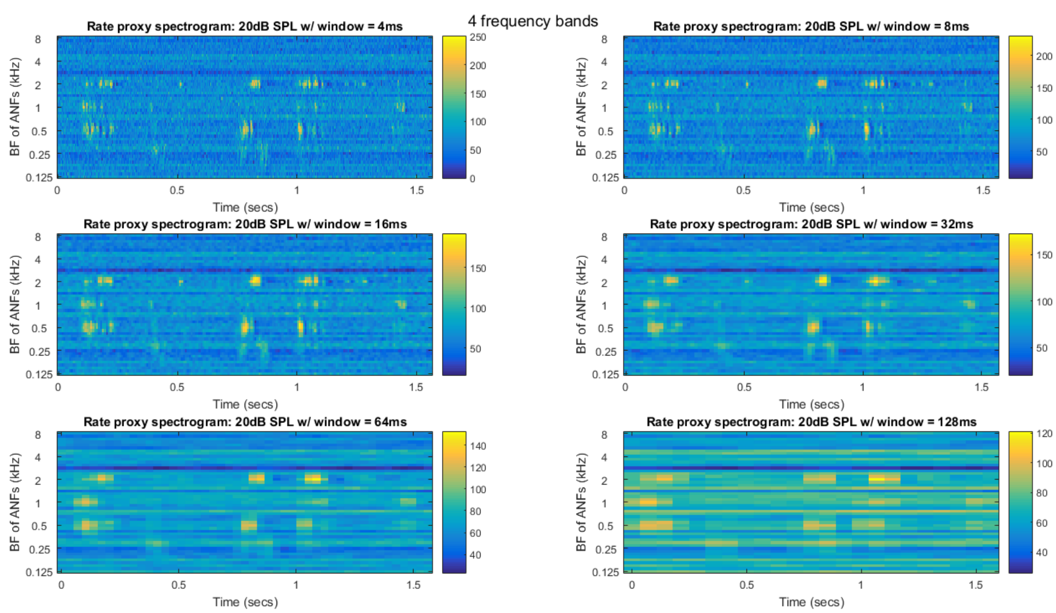

1 frequency bands
Rate proxy spectrogram: 80dB SPL w/ window = 4ms
Rate proxy spectrogram: 80dB SPL w/ window = 8ms
Rate proxy spectrogram: 80dB SPL w/ window = 16ms
Rate proxy spectrogram: 80dB SPL w/ window = 32ms
Rate proxy spectrogram: 80dB SPL w/ window = 64ms
Rate proxy spectrogram: 80dB SPL w/ window = 128ms
BF of ANFs (kHz)
Time (secs)

Fig. 21



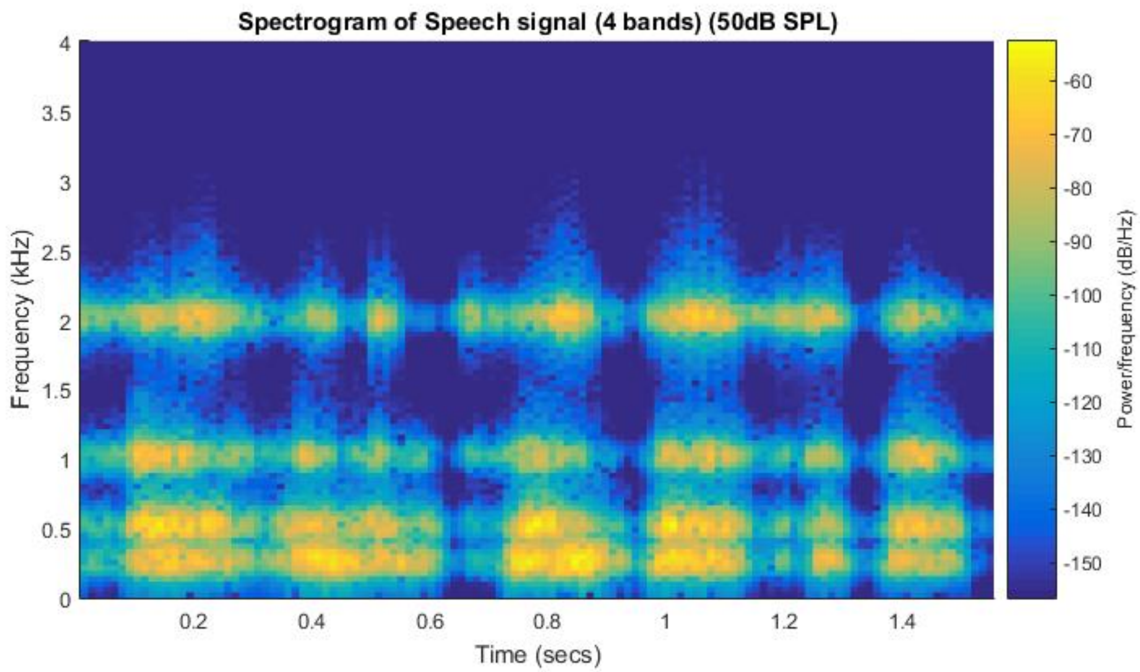
Spectrogram of Speech signal (4 bands) (20dB SPL)
Frequency (kHz)
4
3.5
3
2.5
2
1.5
1
0.5
0
0.2
0.4
0.6
0.8
1
1.2
1.4
Time (secs)
-90
-100
-110
-120
-130
-140
-150
Power/frequency (dB/Hz)

Fig. 22



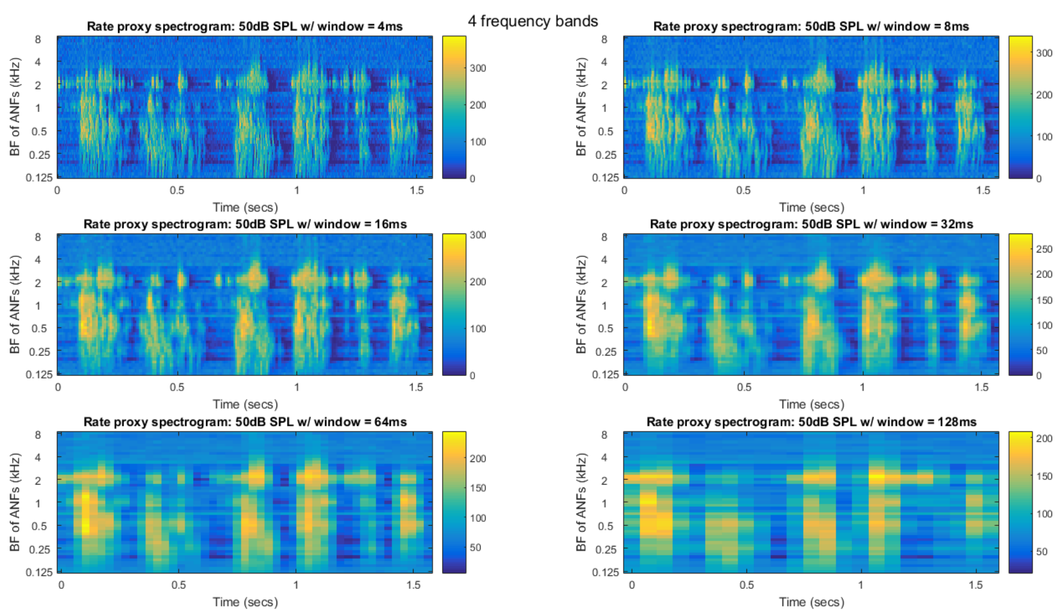
4 frequency bands
Rate proxy spectrogram: 20dB SPL w/ window = 4ms
Rate proxy spectrogram: 20dB SPL w/ window = 8ms
Rate proxy spectrogram: 20dB SPL w/ window = 16ms
Rate proxy spectrogram: 20dB SPL w/ window = 32ms
Rate proxy spectrogram: 20dB SPL w/ window = 64ms
Rate proxy spectrogram: 20dB SPL w/ window = 128ms
BF of ANFs (kHz)
Time (secs)

Fig. 23



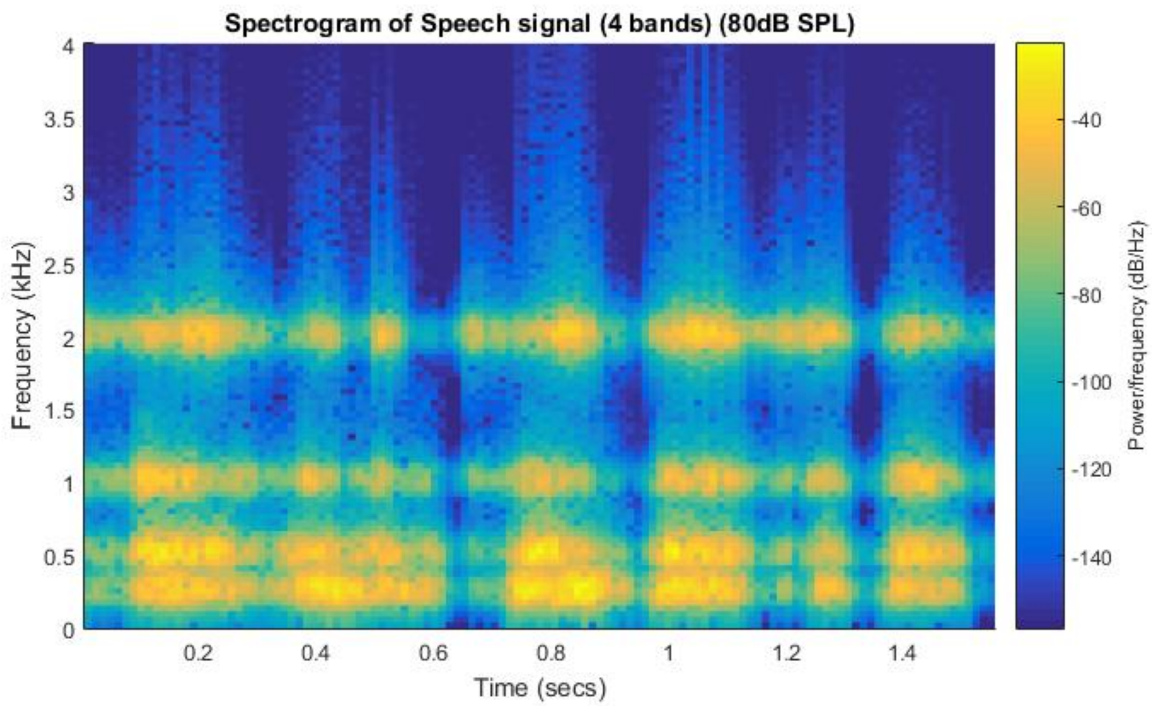
Spectrogram of Speech signal (4 bands) (50dB SPL)
Frequency (kHz)
Time (secs)
Power/frequency (dB/Hz)

Fig. 24



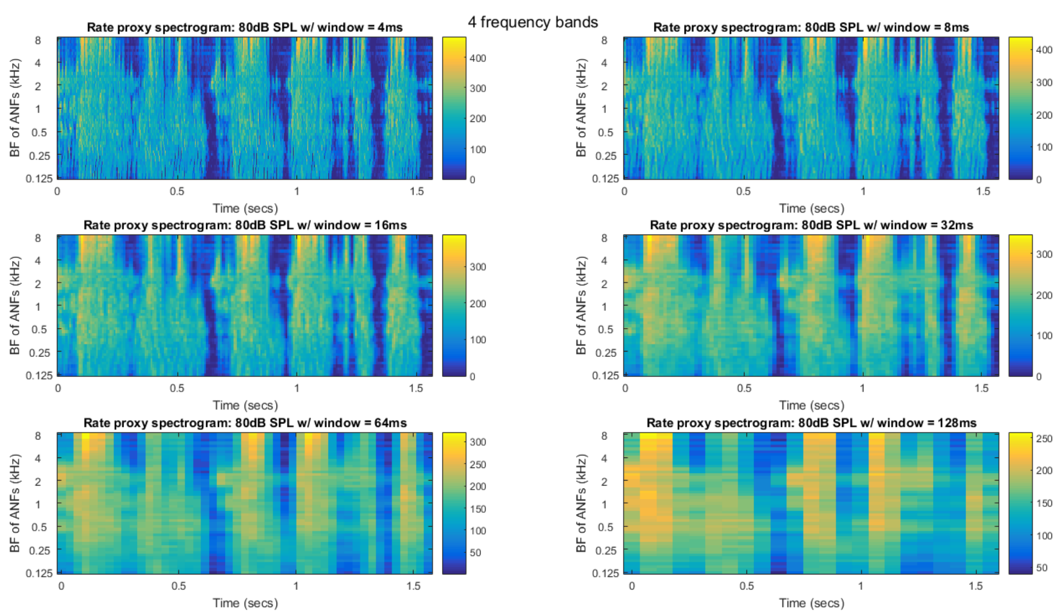
4 frequency bands
Rate proxy spectrogram: 50dB SPL w/ window = 4ms
Rate proxy spectrogram: 50dB SPL w/ window = 8ms
Rate proxy spectrogram: 50dB SPL w/ window = 16ms
Rate proxy spectrogram: 50dB SPL w/ window = 32ms
Rate proxy spectrogram: 50dB SPL w/ window = 64ms
Rate proxy spectrogram: 50dB SPL w/ window = 128ms
BF of ANFs (kHz)
Time (secs)

Fig. 25



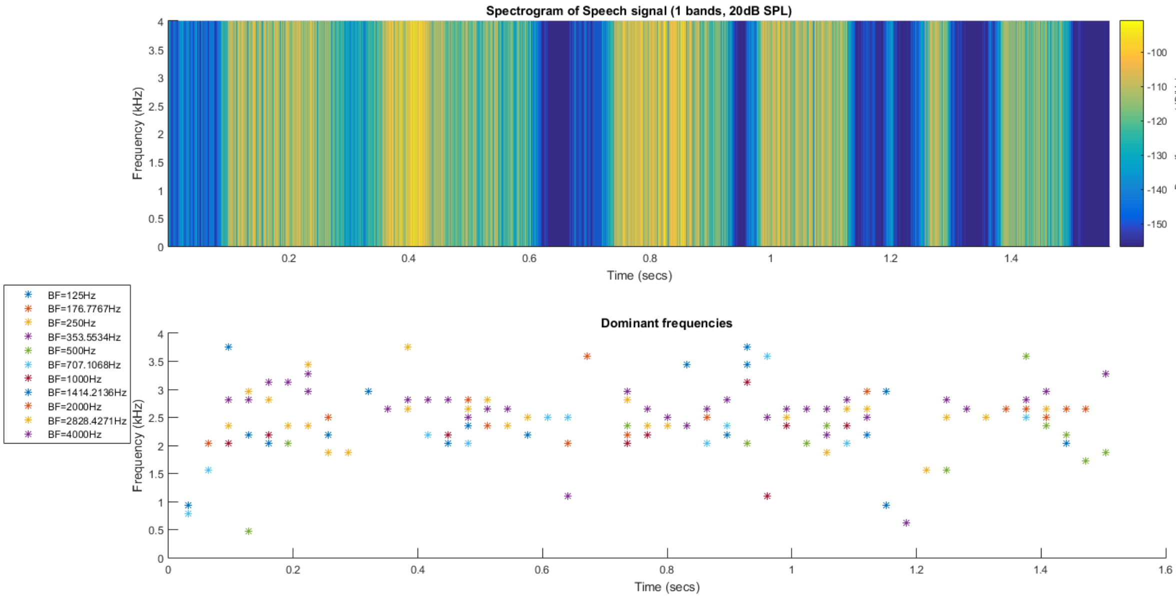
Spectrogram of Speech signal (4 bands) (80dB SPL)
Frequency (kHz)
Time (secs)
Power/frequency (dB/Hz)

Fig. 26



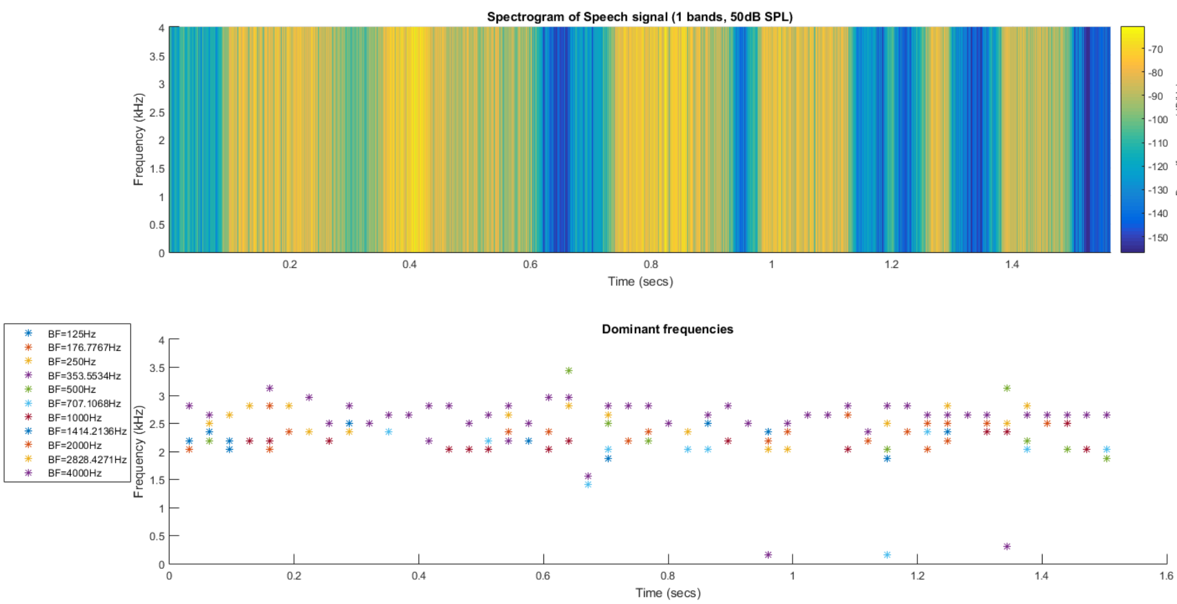
4 frequency bands
Rate proxy spectrogram: 80dB SPL w/ window = 4ms
Rate proxy spectrogram: 80dB SPL w/ window = 8ms
Rate proxy spectrogram: 80dB SPL w/ window = 16ms
Rate proxy spectrogram: 80dB SPL w/ window = 32ms
Rate proxy spectrogram: 80dB SPL w/ window = 64ms
Rate proxy spectrogram: 80dB SPL w/ window = 128ms
BF of ANFs (kHz)
Time (secs)

Fig. 27



Spectrogram of Speech signal (1 bands, 20dB SPL)
Frequency (kHz)
Time (secs)
Power/frequency (dB/Hz)
Dominant frequencies
BF=125Hz
BF=176.7767Hz
BF=250Hz
BF=353.5534Hz
BF=500Hz
BF=707.1068Hz
BF=1000Hz
BF=1414.2136Hz
BF=2000Hz
BF=2828.4271Hz
BF=4000Hz

Fig. 28



Spectrogram of Speech signal (1 bands, 50dB SPL)
Frequency (kHz)
Time (secs)
Power/frequency (dB/Hz)
Dominant frequencies
BF=125Hz
BF=176.7767Hz
BF=250Hz
BF=353.5534Hz
BF=500Hz
BF=707.1068Hz
BF=1000Hz
BF=1414.2136Hz
BF=2000Hz
BF=2828.4271Hz
BF=4000Hz

Fig. 29

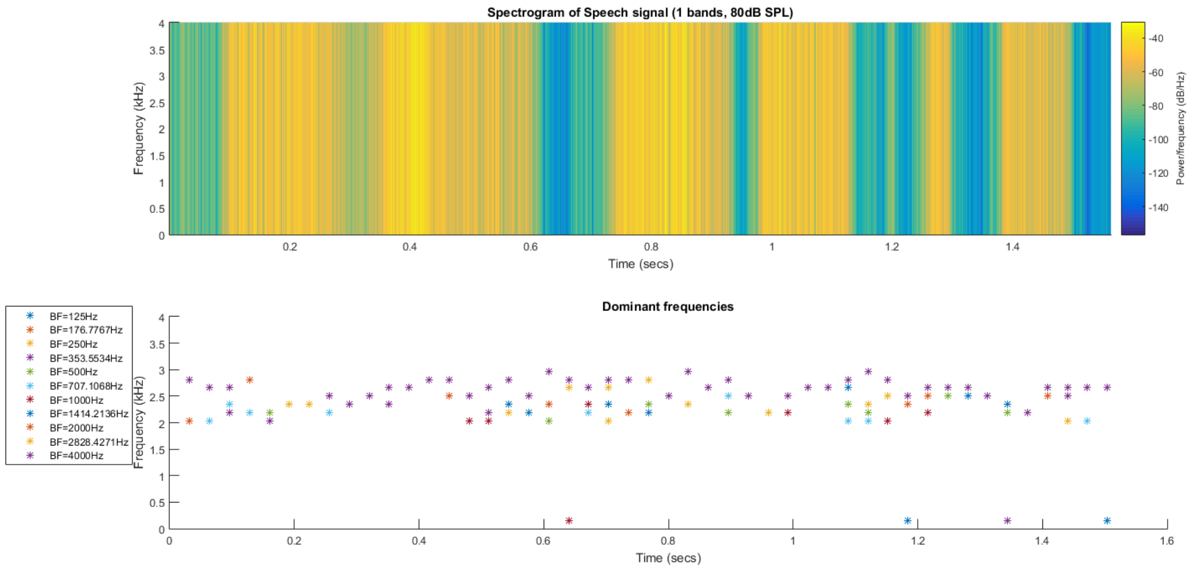

Spectrogram of Speech signal (1 bands, 80dB SPL)
Frequency (kHz)
Time (secs)
Power/frequency (dB/Hz)
Dominant frequencies
BF=125Hz
BF=176.7767Hz
BF=250Hz
BF=353.5534Hz
BF=500Hz
BF=707.1068Hz
BF=1000Hz
BF=1414.2136Hz
BF=2000Hz
BF=2828.4271Hz
BF=4000Hz

Fig. 30

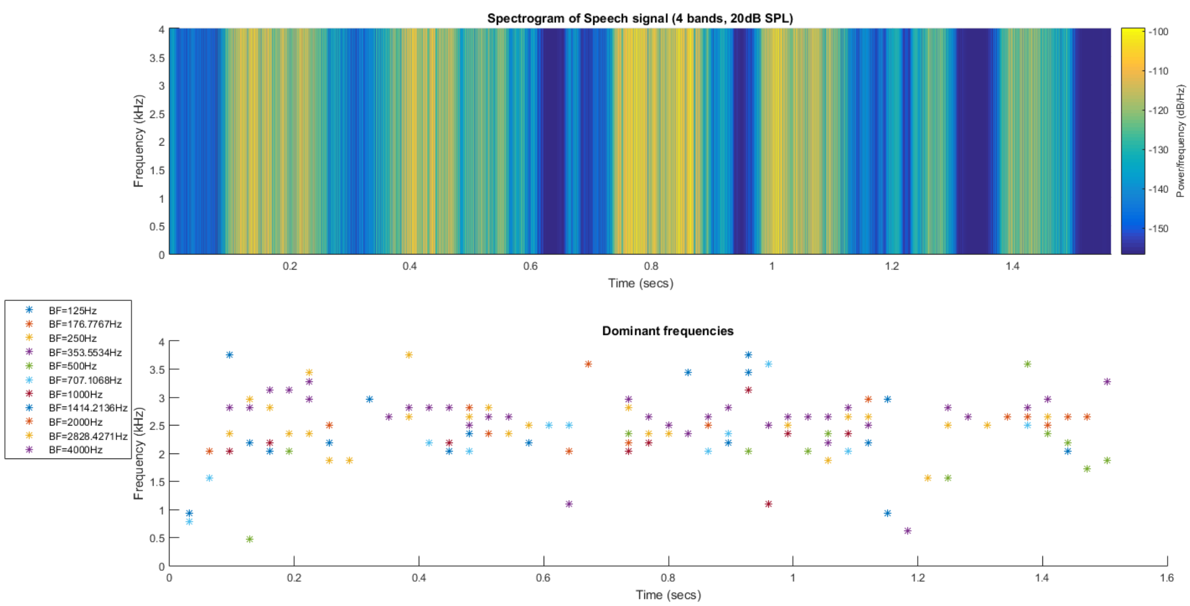

Spectrogram of Speech signal (4 bands, 20dB SPL)
Frequency (kHz)
Time (secs)
Power/frequency (dB/Hz)
Dominant frequencies
BF=125Hz
BF=176.7767Hz
BF=250Hz
BF=353.5534Hz
BF=500Hz
BF=707.1068Hz
BF=1000Hz
BF=1414.2136Hz
BF=2000Hz
BF=2828.4271Hz
BF=4000Hz

Fig. 31

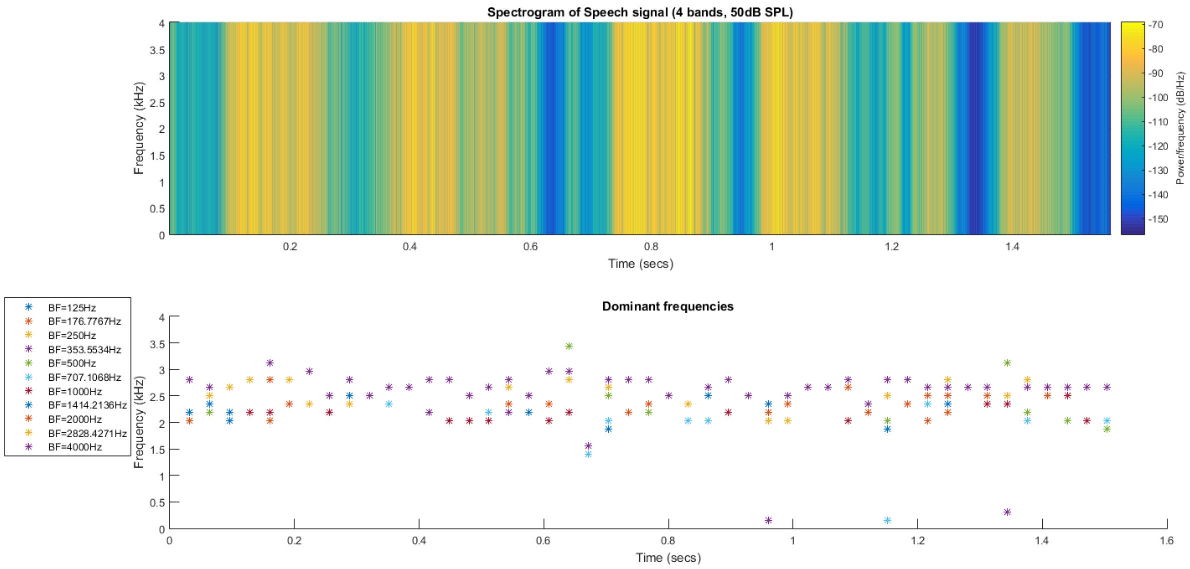

Spectrogram of Speech signal (4 bands, 50dB SPL)
Frequency (kHz)
Time (secs)
Power/frequency (dB/Hz)
Dominant frequencies
Frequency (kHz)
Time (secs)
BF=125Hz
BF=176.7767Hz
BF=250Hz
BF=353.5534Hz
BF=500Hz
BF=707.1068Hz
BF=1000Hz
BF=1414.2136Hz
BF=2000Hz
BF=2828.4271Hz
BF=4000Hz

Fig. 32

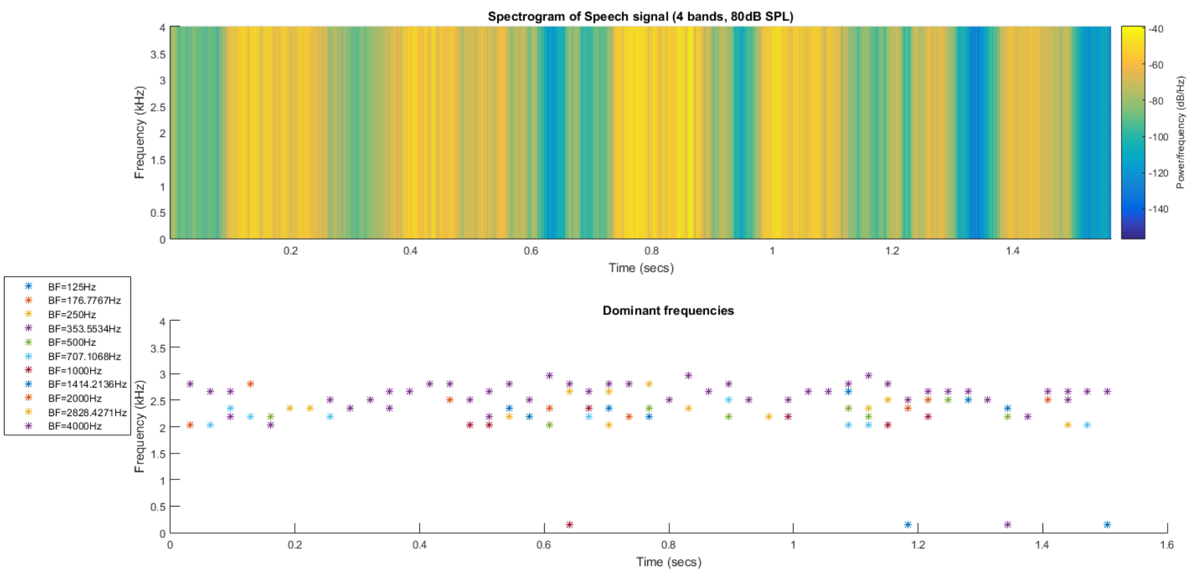

Spectrogram of Speech signal (4 bands, 80dB SPL)
Frequency (kHz)
Time (secs)
Power/frequency (dB/Hz)
Dominant frequencies
Frequency (kHz)
Time (secs)
BF=125Hz
BF=176.7767Hz
BF=250Hz
BF=353.5534Hz
BF=500Hz
BF=707.1068Hz
BF=1000Hz
BF=1414.2136Hz
BF=2000Hz
BF=2828.4271Hz
BF=4000Hz

Fig. 33