ECE 438 Digital Signal Processing
Week 12: Speech Processing (Lab 9a)

Date 4/15/2020

Section 2

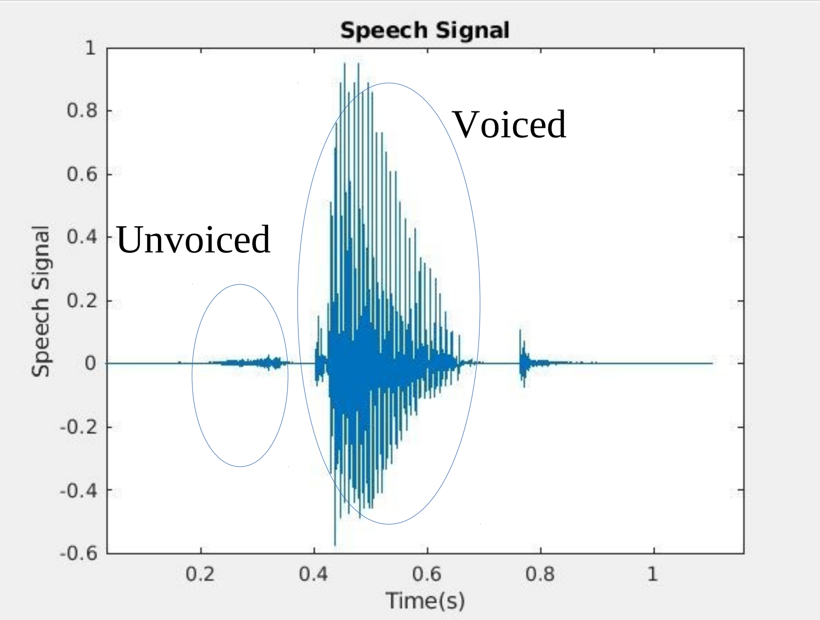
Name		Sign	Time spent outside lab
David Dang	[%]	David Dang	24
Benedict Lee	[%]	Benedict Lee	24

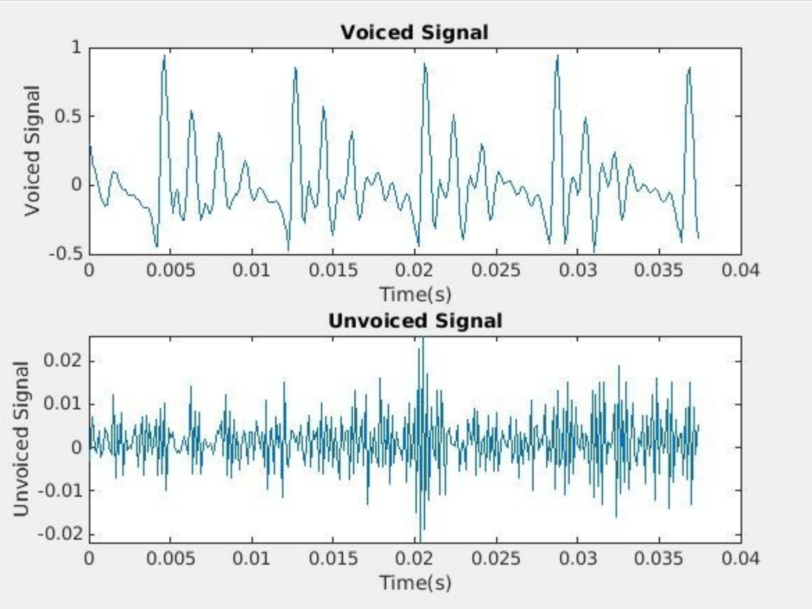
Grading Rubric (Spring 2020)

	below expectations	lacks in some respect	meets all expectations
Completeness of the report			
Organization of the report *One-sided, with cover sheet, answers are in the same order as questions in the lab, copies of the questions*			
Quality of figures *Correctly labeled with title, x-axis, y-axis, and name(s)*			
Understanding differences between voiced/unvoiced segments(30 pts) *Matlab plots and code(zero_cross), questions*			
Understanding and implementation of short-time DTFT (30 pts) *Matlab plots and code(DFTwin), questions*			
Understanding and implementation of spectrogram (40 pts) *Matlab plots and code(specgm), questions, formant estimates*			

ECE 438 Lab 9a
David Dang & Benedict Lee

2.2 Classification of Voiced/Unvoiced Speech
INLAB REPORT: Hand in your labeled plots. Explain how you selected your voiced and unvoiced regions.





- The voiced region is the part of the waveform that is quasi-periodic and the unvoiced region is the part of the waveform that has a lot of zero-crossings

INLAB REPORT:
1. Give your estimate of the pitch period for the voiced segment, and your prediction of the gender of the speaker.
2. For each of the two vectors,VoicedSig and UnvoicedSig, list the average energy and number of zero-crossings. Which segment has a greater average energy? Which segment has a greater zero-crossing rate?
3. Hand in your zero cross function.

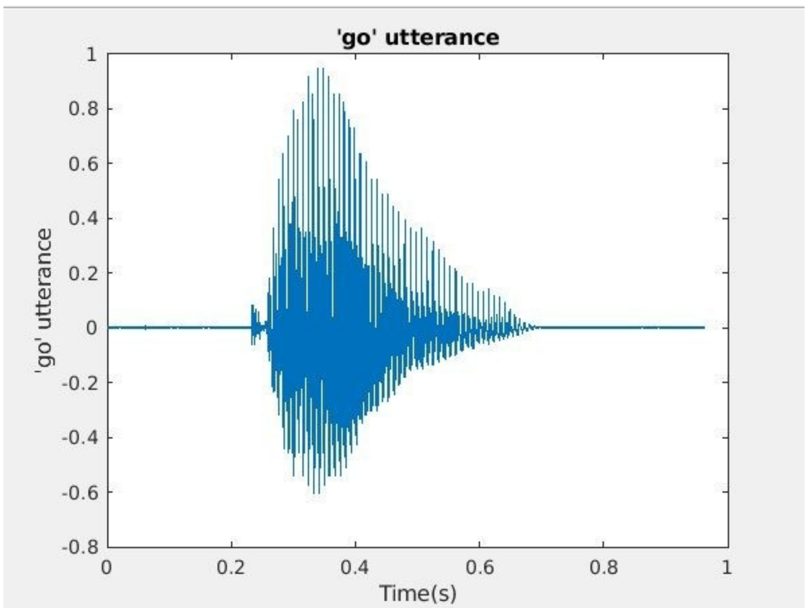
- From the Voiced Signal graph, the time difference between the 2 peaks at 0.020625s and 0.02875s is 8.125 ms, which is the pitch period for the voiced segment. Based on this pitch period, the gender of the speaker is male.
- The average energy for the voiced segment is 0.0693 and 5.3152e-05 for the unvoiced segment. The number of zero crossings for the voiced segment is 53, while that for the unvoiced segment is 219; hence, the unvoiced segment has a greater zero crossing rate, while the voiced segment has a greater average energy.

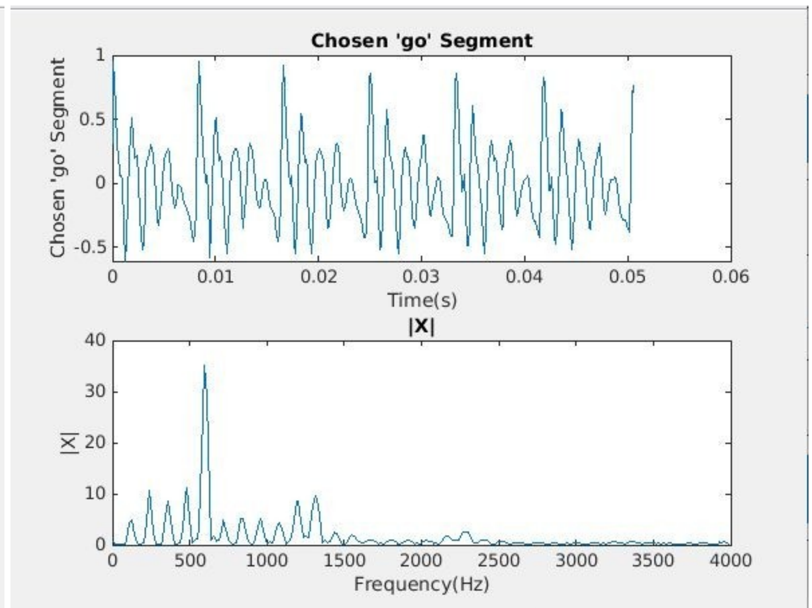
```
function zcdOut = zero_cross(vector)
zcd = dsp.ZeroCrossingDetector;
zcdOut = zcd(vector);
```

3.1 stDTFT

INLAB REPORT: Hand in the code for your DFTwin() function, and your plot. Describe the general shape of the spectrum, and estimate the formant frequencies for the region of voiced speech.

```
function X = DFTwin(x,L,m,N)
w = hamming(L);
x = x(1+m:m+L);
x = x.*w;
X = fft(x, N);
```





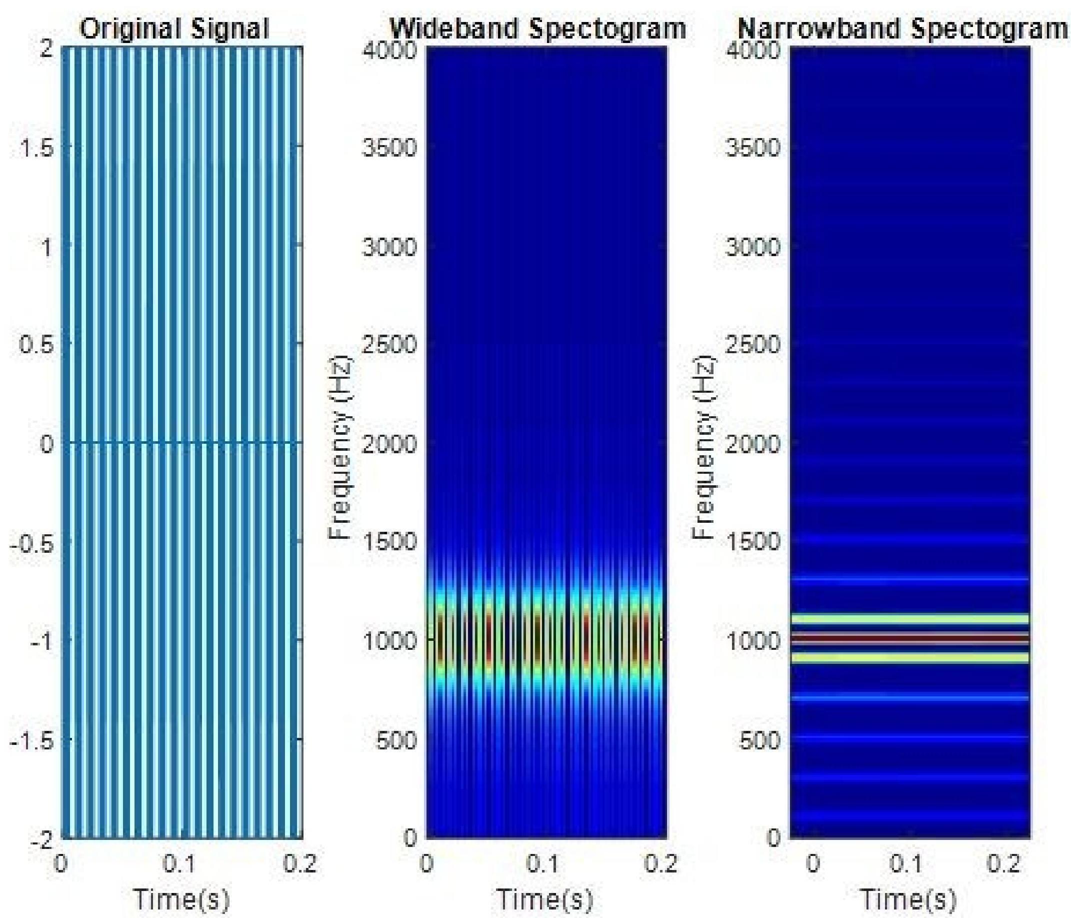
- The spectrum has multiple peaks, with a peak-of-peaks at about 570 Hz surrounded by several other smaller peaks. Further away from this frequency, the magnitude of the spectrum decays to 0. From the graph, the formant frequencies of significance are estimated to be at 570 Hz, 495 Hz, 400 Hz, 250 Hz, 1200 Hz and 1300 Hz.

3.2 The Spectrogram

INLAB REPORT: Hand in your code for Specgm() and your plots. Do you see vertical striations in the wideband spectrogram? Similarly, do you see horizontal striations in the narrowband spectrogram? In each case, what causes these lines, and what does the spacing between them represent?

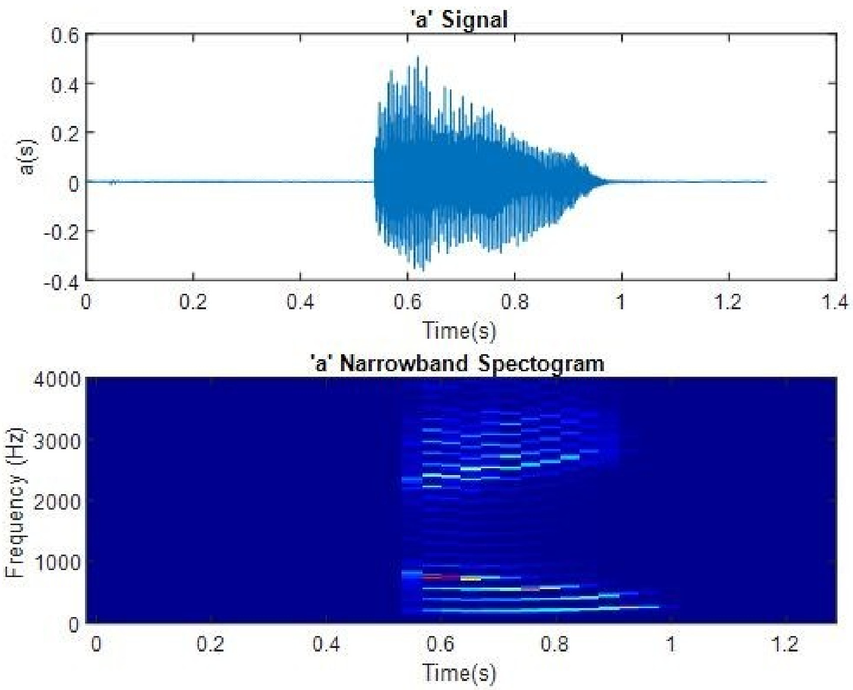
```
function A = Specgm(x, L, overlap, N)
m = 0;
n = 1;
B = zeros(N,1);
while (m+L)<=length(x)
    B(1:N) = DFTwin(x,L,m,N);
    A(1:N/2,n) = abs(B(1:N/2));
    m = m+L-overlap;
    n = n + 1;
end
imagesc([0 length(x)/8000], [0 4000], A)
axis xy
colormap(jet)
```

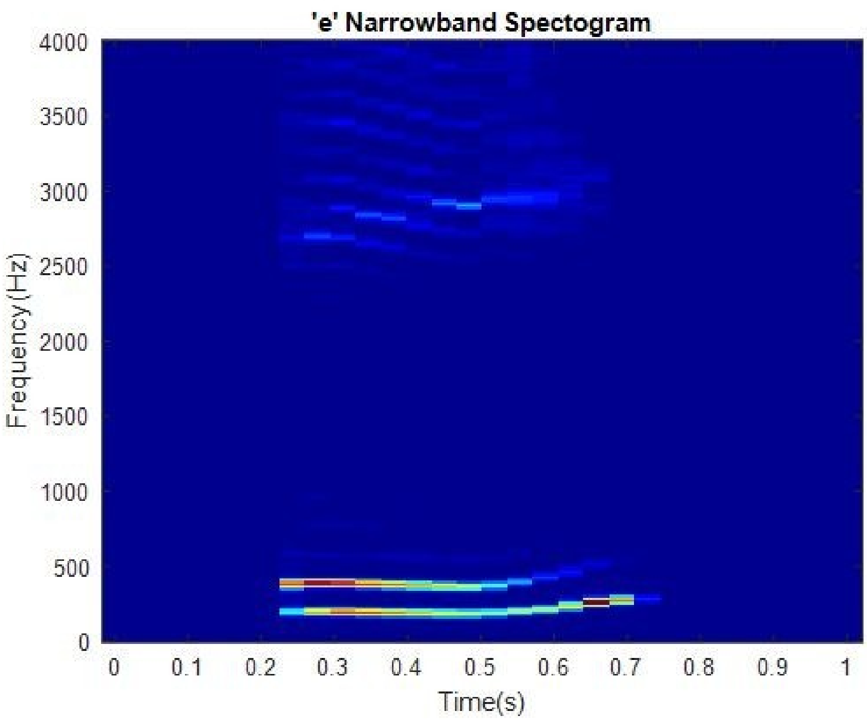
- Vertical striations are present in the wideband spectrogram, whereas horizontal strations are present in the narrowband spectrogram. The lines in each spectrogram represent a single DTFT, and the spacing between them is representative of the size of the window.

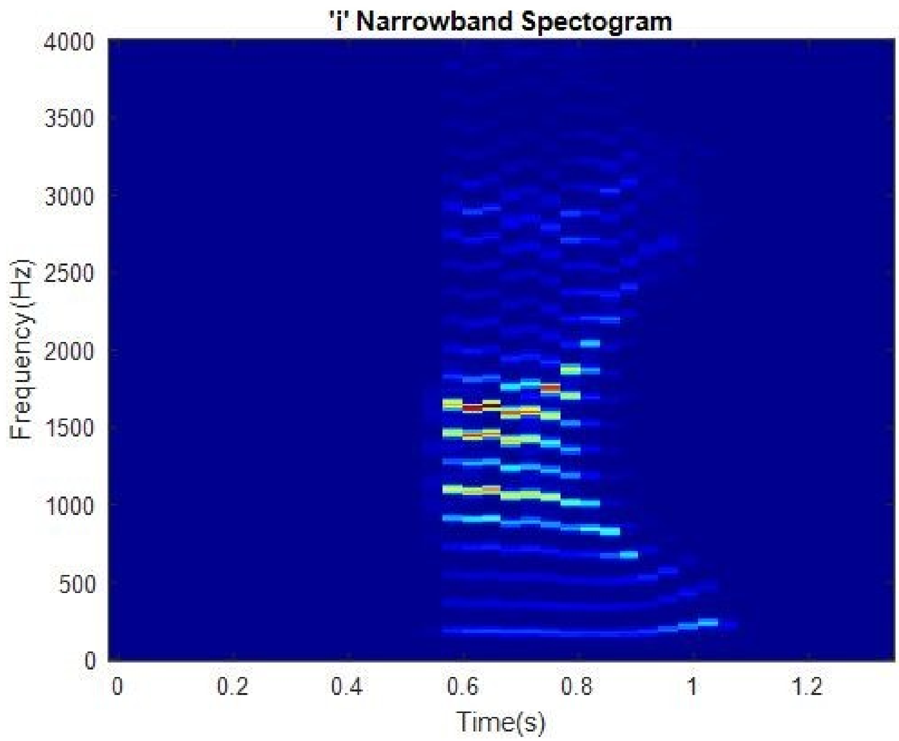


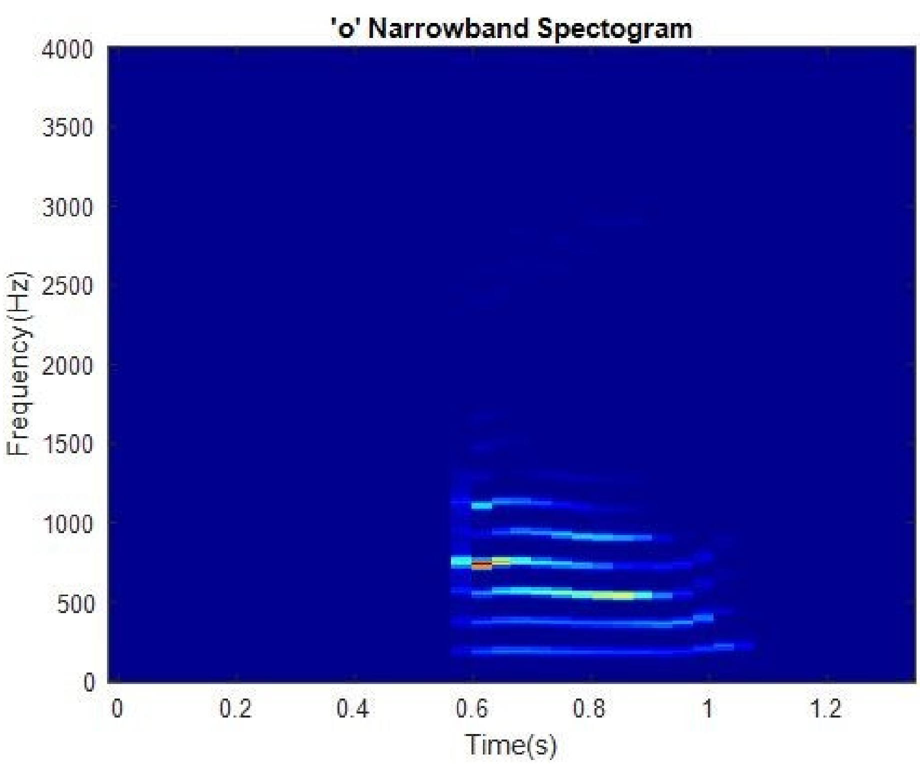
3.3 Formant Analysis

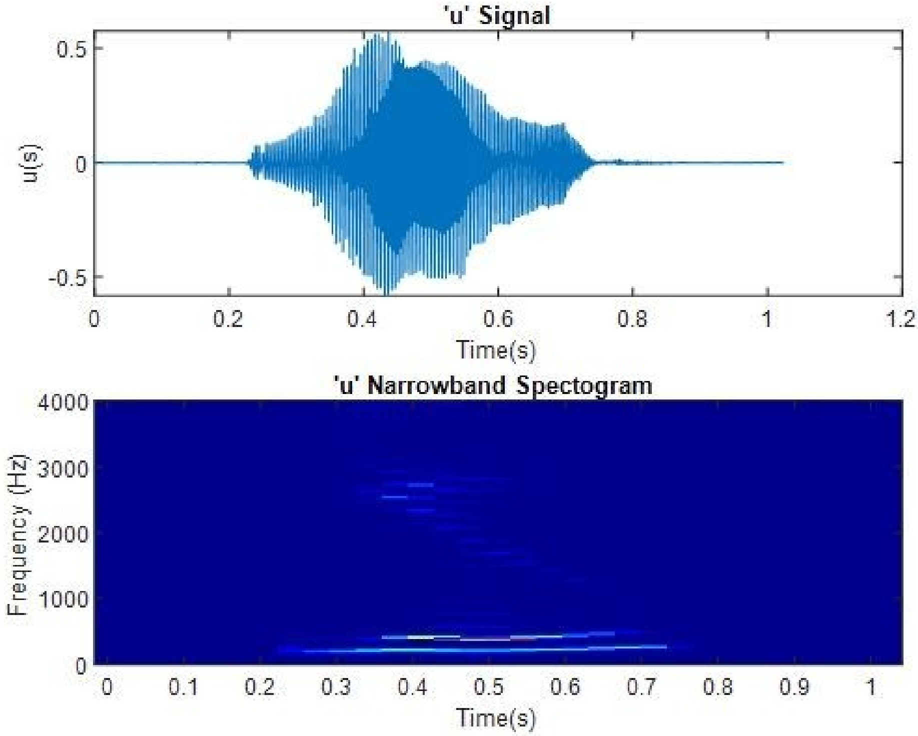
INLAB REPORT: Hand in your formant estimates on the vowel triangle.

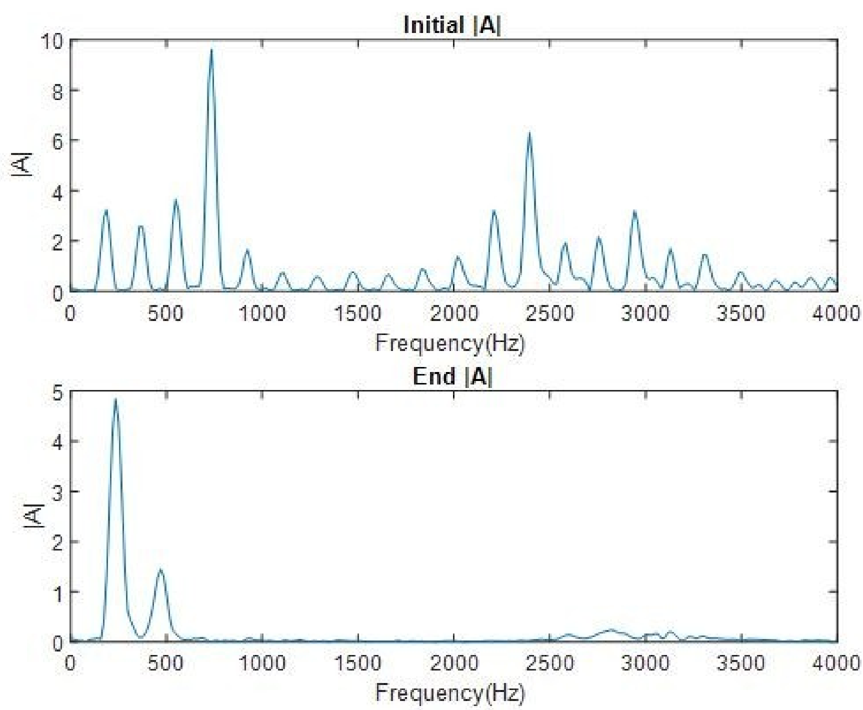




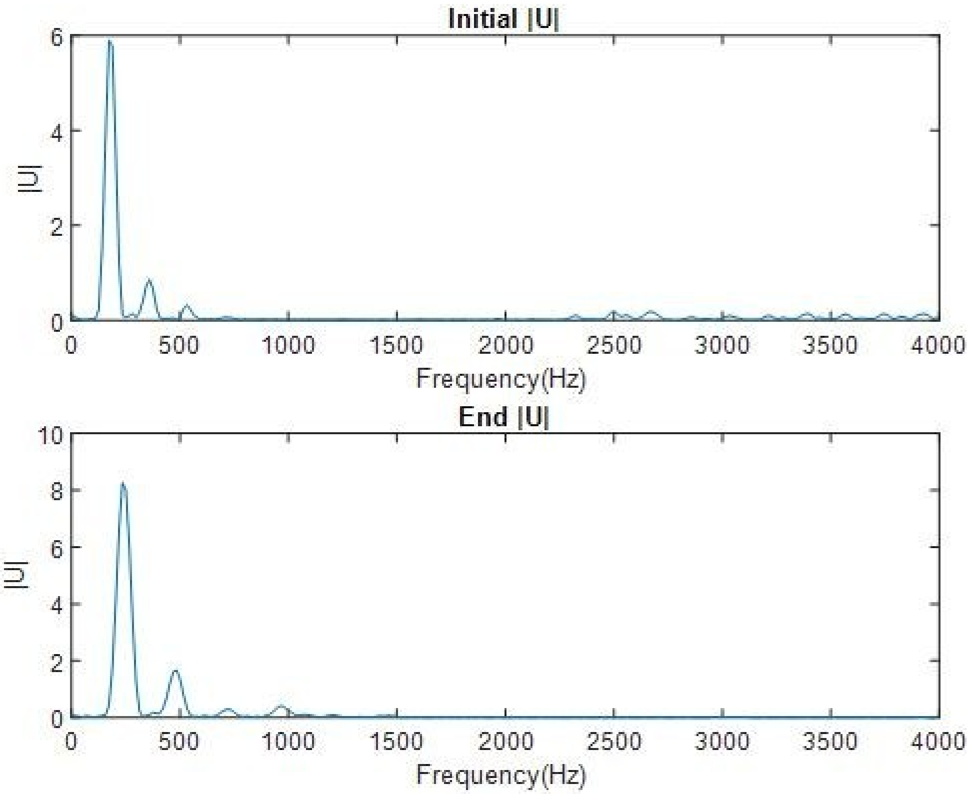

'i' Narrowband Spectogram
4000
3500
3000
2500
2000
1500
1000
500
0
Frequency (Hz)
0
0.2
0.4
0.6
0.8
1
1.2
Time(s)


'o' Narrowband Spectogram
4000
3500
3000
2500
2000
1500
1000
500
0
Frequency (Hz)
0
0.2
0.4
0.6
0.8
1
1.2
Time(s)





- From the initial |A| plot, the first formant frequency at the start of the ‘a’ utterance occurs at 735.8 Hz, while the second occurs at 2395 Hz. From the end |A| plot, the first formant frequency at the end of the ‘a’ utterance occurs at 234.8 Hz, while the second occurs at 469.7 Hz.



- From the initial |U| plot, the first formant frequency at the start of the ‘u’ utterance occurs at 172.2 Hz, while the second occurs at 360.1 Hz. From the end |U| plot, the first formant frequency at the end of the ‘u’ utterance occurs at 234.8 Hz, while the second occurs at 485.3 Hz.

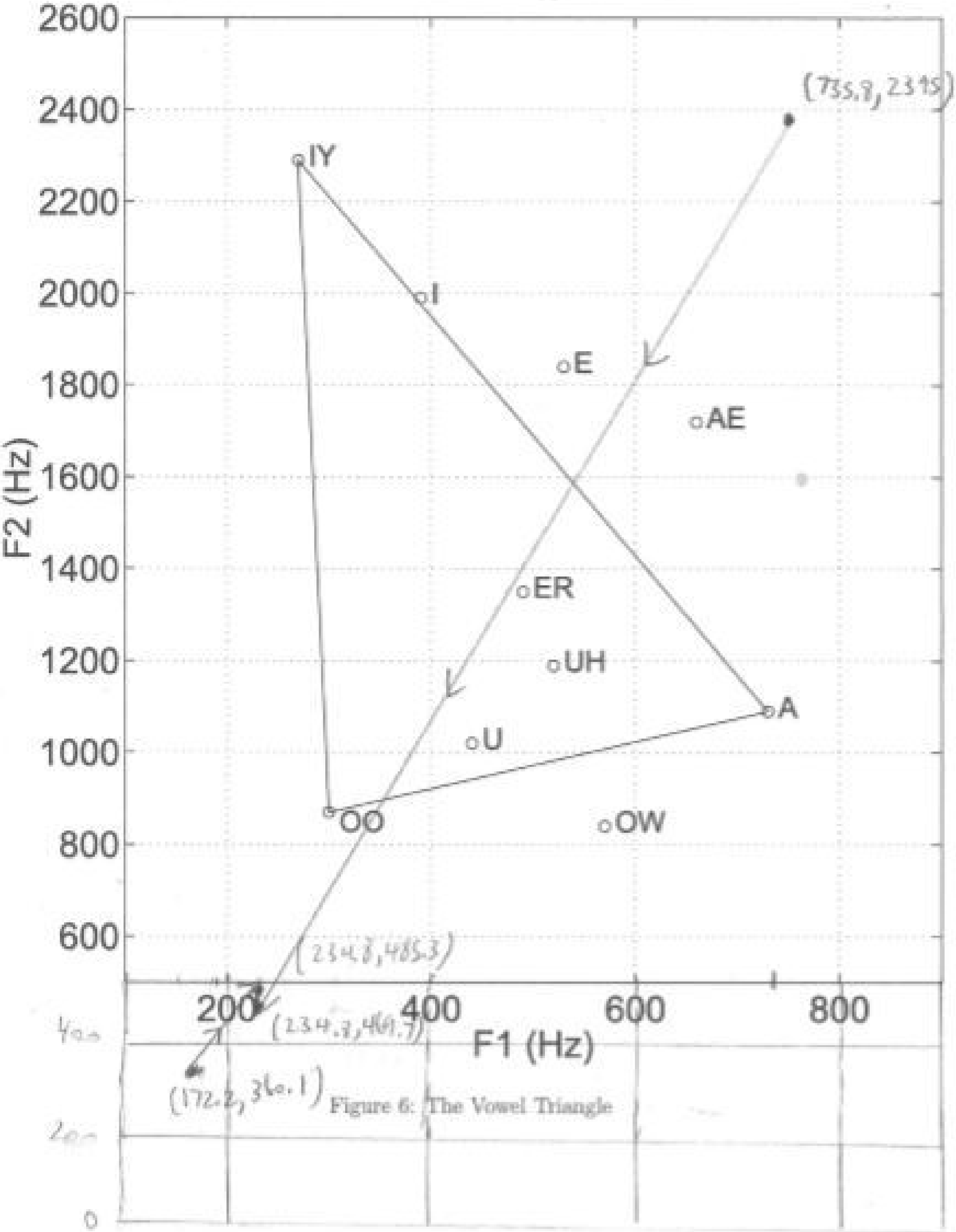


Figure 6: The Vowel Triangle