# Lab 22 Fast Fourier Transform of Human Voice

Bill Hung
17508938
EE145M

Lab Time: 9-12pm Wednesday
Lab Partner: Chih-Chieh Wang (Dennis)
Station 8

## Aim

To sample Human voice periodically when various vowel sounds were pronounced. Fast Fourier Transform (FFT) changes the sampled data from time domain to frequency domain. To investigate features of vowels and features of different speakers.

## 1. Setup

Anti-aliasing Filter – LF356

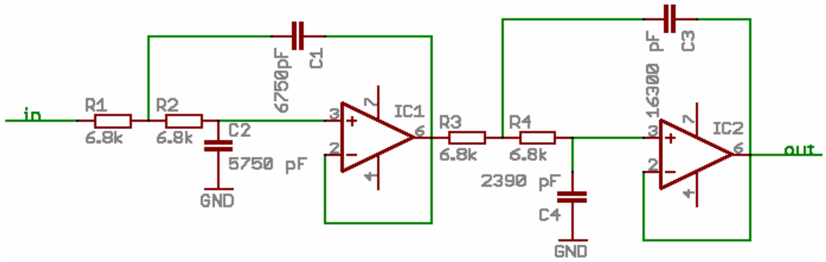


Figure 1. Butterworth four-pole low-pass filter. The filter reduces the frequency magnitude at and above one-half of the sampling frequency. In this case, the filter cuts off at about 10kHz. At 7kHz, the amplitude was reduced to 0.707 (the corner frequency of the filter), and preserve the amplitude at frequencies below 7kHz (gain=1). In this lab, we tried to construct this anti-aliasing filter, but it was not successful. The data were sampled without the anti-aliasing filter in this lab.

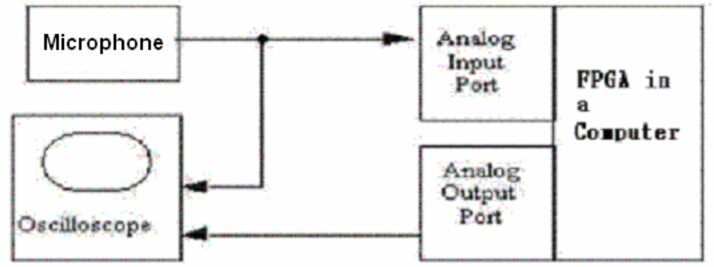


Figure 2. Setup Diagram for Lab 22. The microphone was supposed to have 500mV to 1V peak-to-peak voltage, the actual peak-to-peak voltage was about 250mV.

## 2. Data summary

### 2.1 Plot the time samples of the four data sets (procedure section 6) to compare the time waveforms of different vowels spoken by the same speaker and the same vowels spoken by different speakers.

Sampling time = 2048 samples * (1/20kHz) = 0.1024s

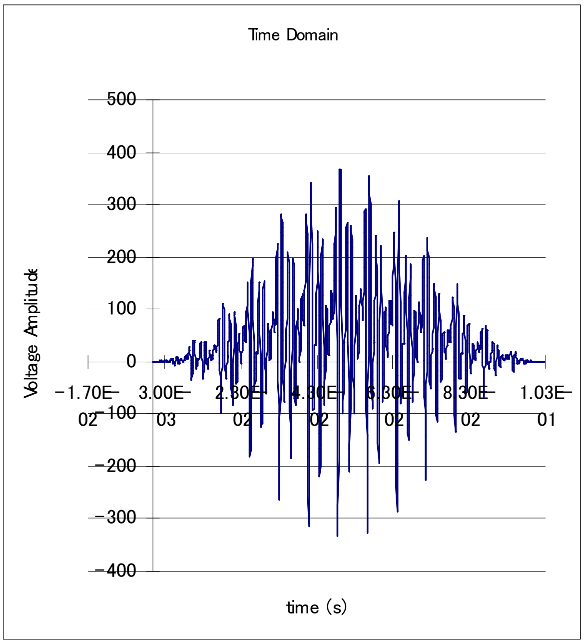


**Figure 3 Vowel "a" by Bill Time Plot**

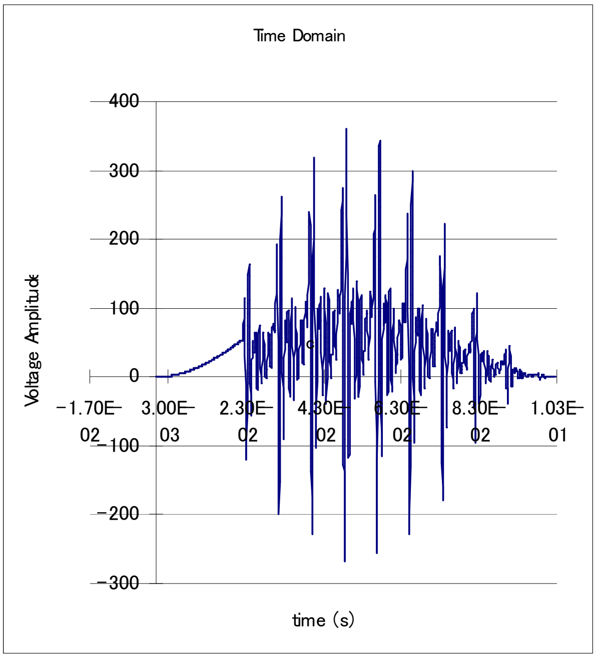


**Figure 4 Vowel "a" by Dennis Time Plot**

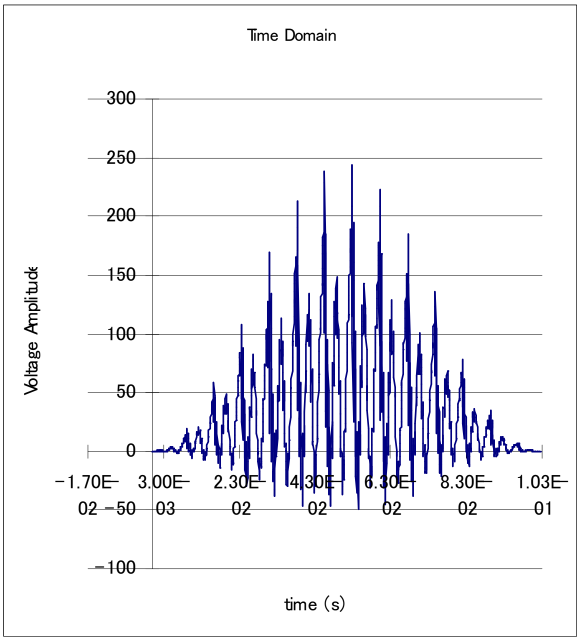


**Figure 5 Vowel "i" by Bill Time Plot**

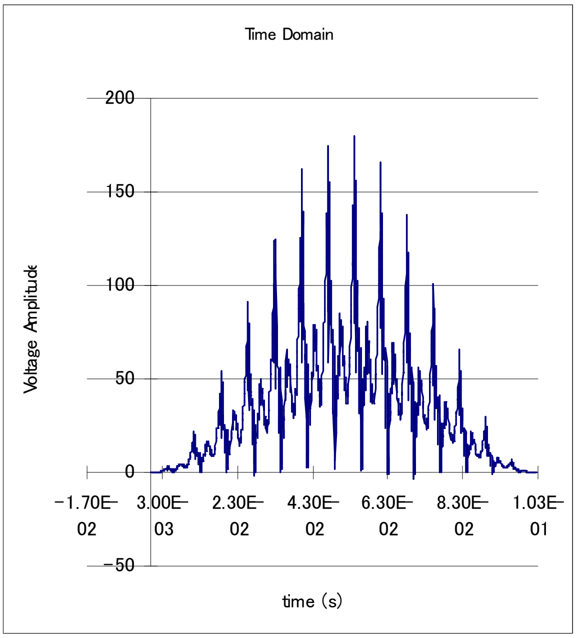


**Figure 6 Vowel "i" by Dennis Time Plot**

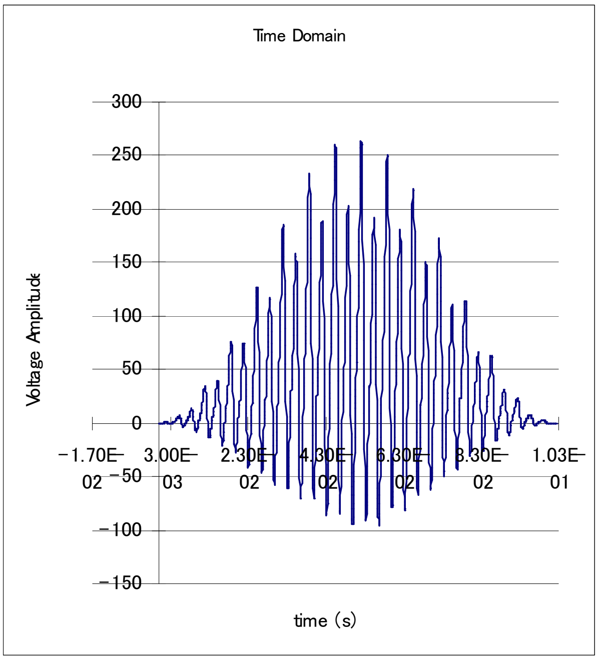


**Figure 7 Vowel "u" by Bill Time Plot**

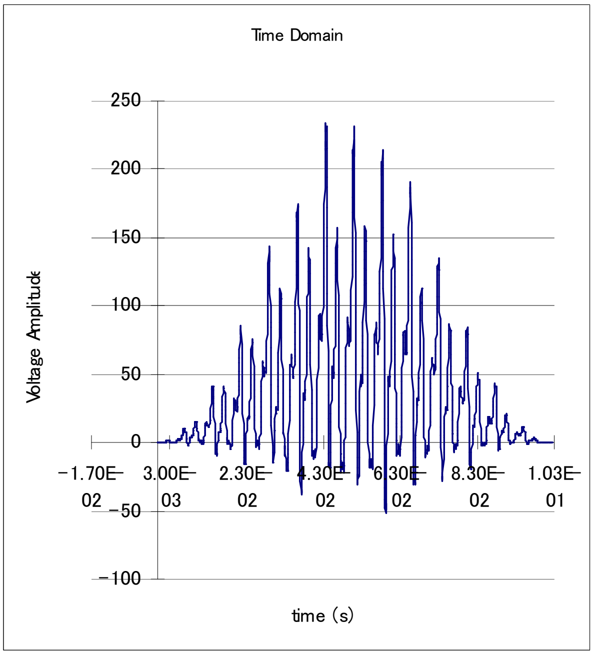


**Figure 8 Vowel "u" by Dennis Time Plot**

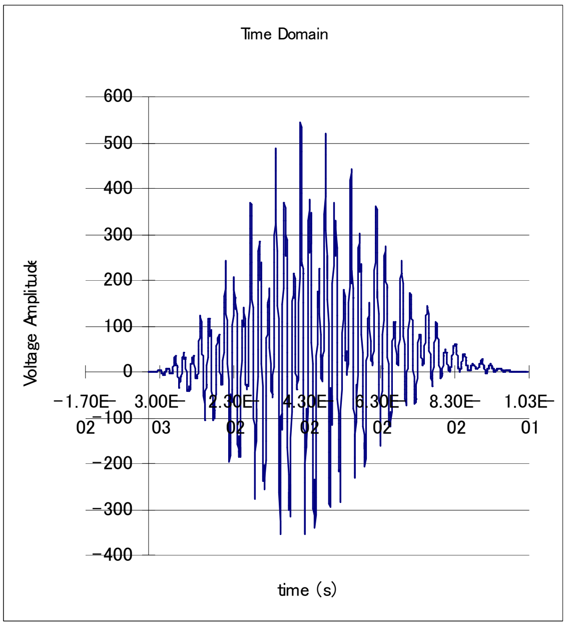


**Figure 9 Vowel "er" by Bill Time Plot**

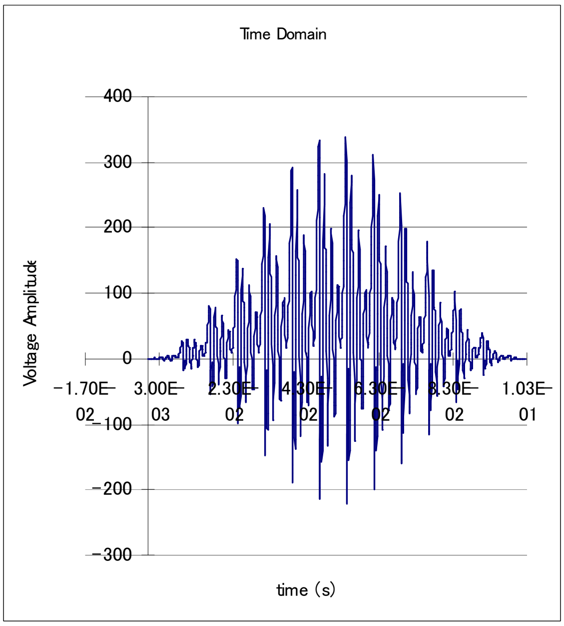


**Figure 10 Vowel "er" by Dennis Time Plot**

Overall, the time-domain plots are very similar, and it is difficult to distinguish which vowel was spoken from the time-domain plot by the same speaker. The plots are generated from the FPGA data, and original time-domain data was filtered by a Hann window. The Hann windows reduces the time-domain signal gradually at the two ends of the time-domain plots, which would reduce spectral leakage in the frequency domain.

There are differences in the amplitudes of the waveforms. The amplitude represents the loudness of the sound, which give irrelevant information about which vowel was spoken. Therefore, the amplitudes of the waveforms were not compared.

By visual inspections, the waveforms between two speakers are similar, and it is ineffective to identify individual frequency components from the time-domain plots. After filtering by the Hann window the time-domain waveforms are even harder to distinguish, so the frequency plot is preferred to analyze the frequency components of the waveforms.

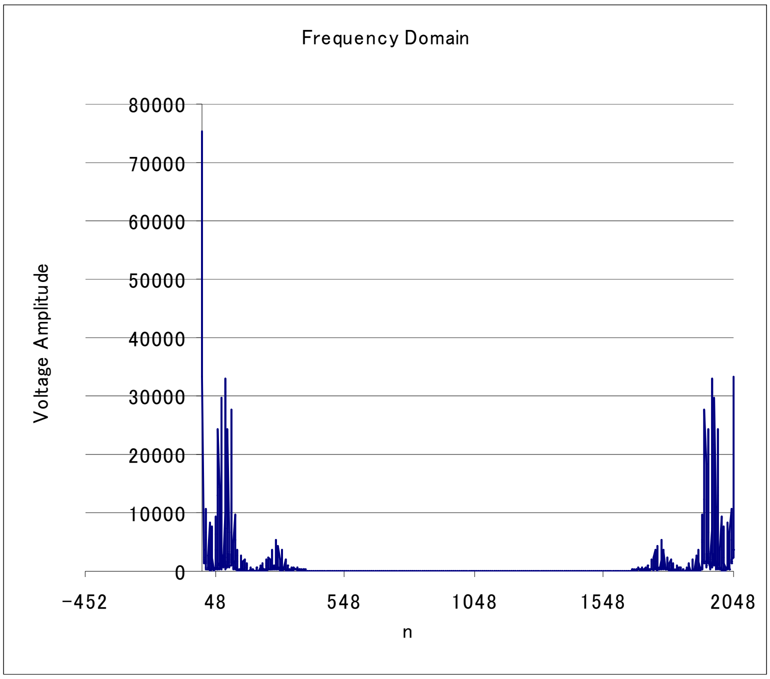


**Figure 11 Frequency Plot after the FFT**

This is an example of a waveform in the frequency domain after the Fast Fourier Transformation. 2048 values of magnitudes were calculated by the real and imaginary components of the Fourier Coefficients. For the analysis of this report, the left part of the plot would be enlarged and the x-axis will be in Hz instead of n.

**2.2 Plot the four set of Fourier magnitudes (procedure section 6) to compare the frequency content of different vowels spoken by the same speaker and the same vowels spoken by different speakers. Include a plot of the first 100 Fourier magnitudes so that the fundamental harmonics of the vocal chords can be seen and measured.**

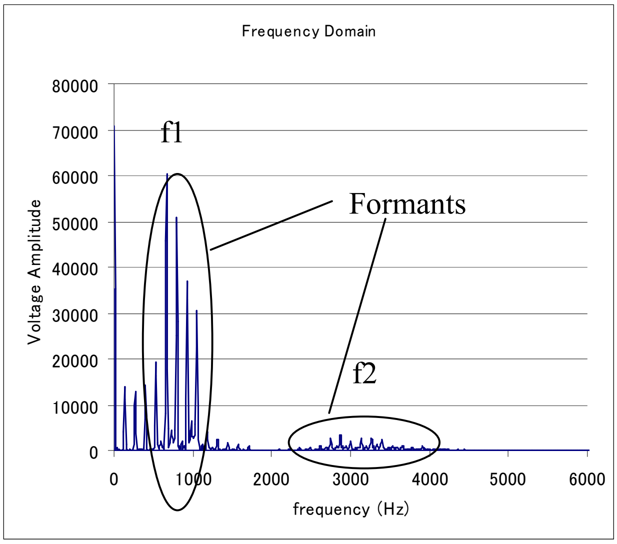


**Figure 12 Vowel “a” by Bill Frequency Plot**

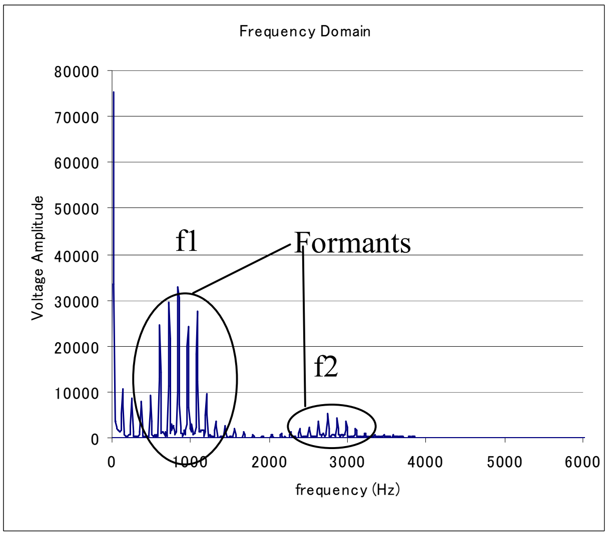


**Figure 13 Vowel “a” by Dennis Frequency Plot**

The magnitudes of the third formant (f3) were too small that f3 could not be determined from the FFT of the sampled data.

For Bill,

| f1(Hz) | f2(Hz) |
|---|---|
| 849.6 | 2753.9 |

For Dennis,

| f1(Hz) | f2(Hz) |
|---|---|
| 664.1 | 2880.9 |

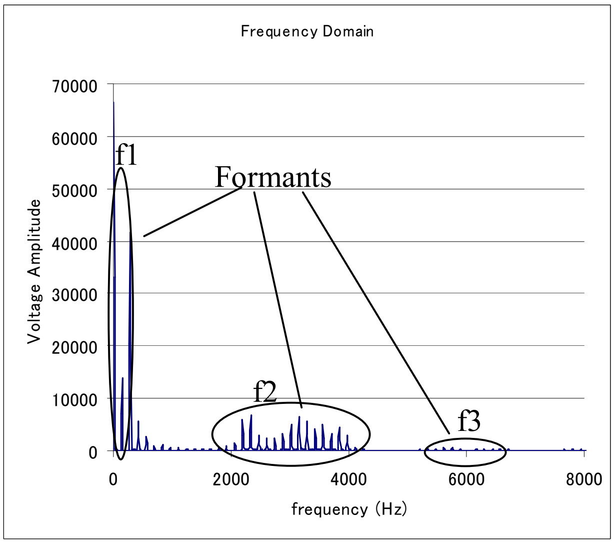


**Figure 14 Vowel “i” by Bill Frequency Plot**

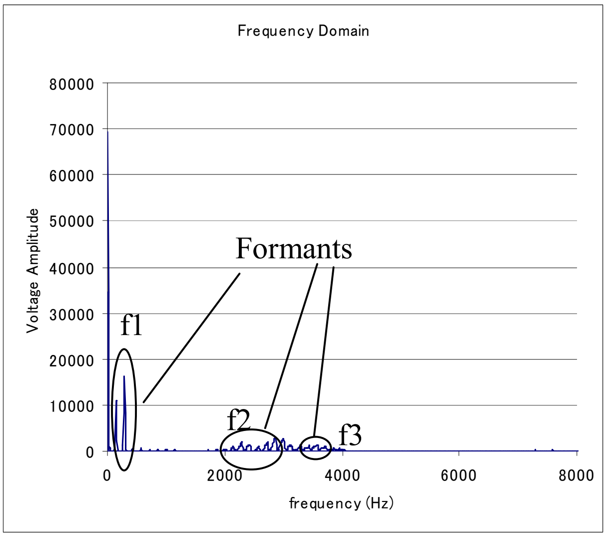


**Figure 15 Vowel “i” by Dennis Frequency Plot**

For Bill,

| f1(Hz) | f2(Hz) | f3(Hz) |
|---|---|---|
| 283.20 | 2334 | 3154.3 |

For Dennis,

| f1(Hz) | f2(Hz) | f3(Hz) |
|---|---|---|
| 293 | 2304.7 | 3017.58 |

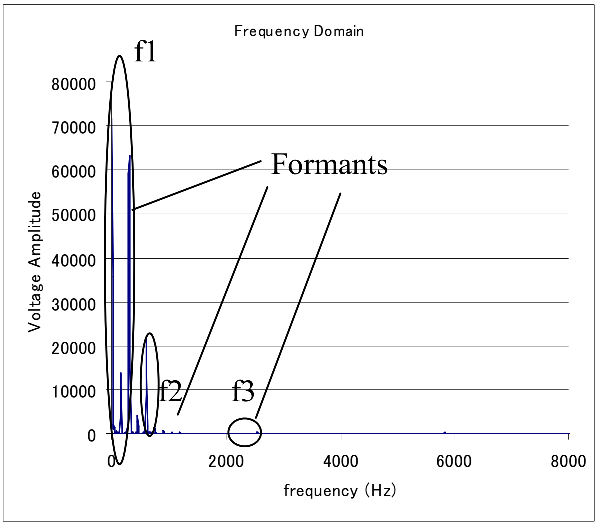


**Figure 16 Vowel "u" by Bill Frequency Plot**

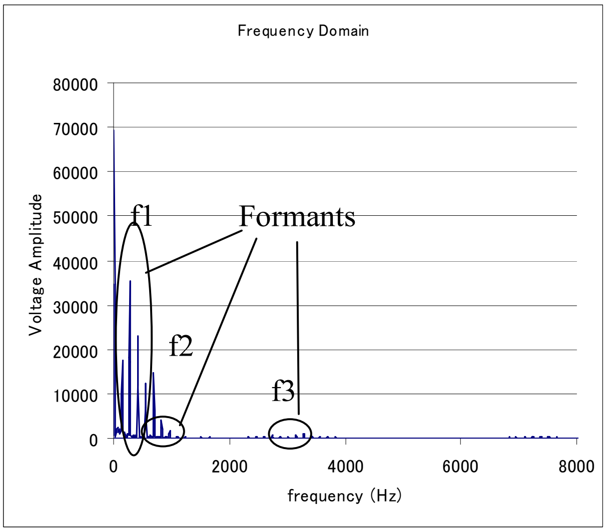


**Figure 17 Vowel "u" by Dennis Frequency Plot**

For Bill,

| f1(Hz) | f2(Hz) | f3(Hz) |
|---|---|---|
| 312.50 | 605.5 | 2627 |

For Dennis,

| f1(Hz) | f2(Hz) | f3(Hz) |
|---|---|---|
| 283.20 | 693.4 | 3339.9 |

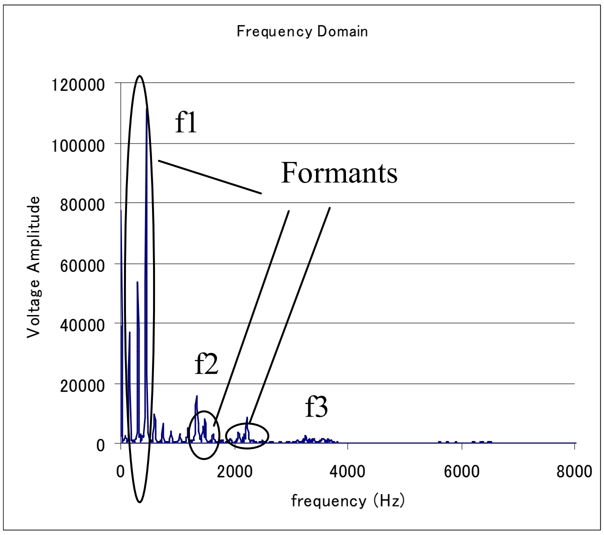


**Figure 18 Vowel "er" by Bill Frequency Plot**

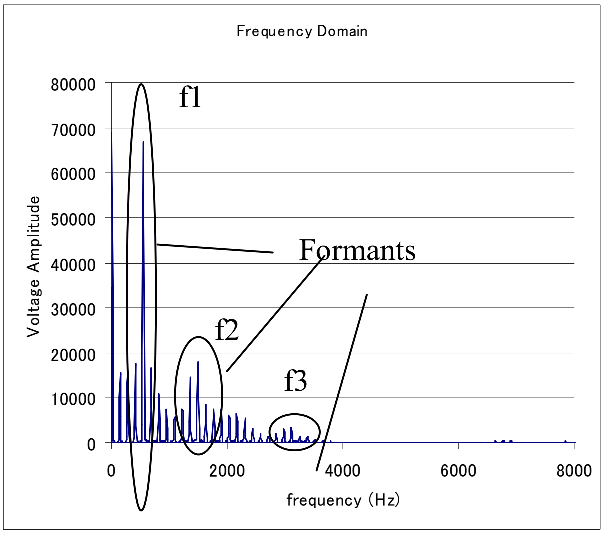


**Figure 19 Vowel "er" by Dennis Frequency Plot**

For Bill,

| f1(Hz) | f2(Hz) | f3(Hz) |
|---|---|---|
| 449.20 | 1337.9 | 2226.6 |

For Dennis,

| f1(Hz) | f2(Hz) | F3(Hz) |
|---|---|---|
| 546.90 | 1494.1 | 3115.2 |

The harmonic spikes are labeled as f1, f2, and f3 respectively. The formants are circled in the plots. Formants are groups of frequencies near the harmonics. The fundamental harmonics measured and indicated by the tables below each of the frequency plots. The first 820 Fourier Amplitudes were plotted (instead of showing just the first 100 Fourier amplitudes), and the harmonics were measured accordingly. The fundamental frequencies and its harmonics are used to calculate the frequency ratio in Section 2.5. The frequency ratio would be used to determine which vowels had been spoken.

**2.3 The vocal chords produce a periodic oscillation that will appear in your Fourier magnitudes as a low-frequency spike (the fundamental) plus all higher harmonics. This is similar to what you observed in Laboratory Exercise 21 for the square wave, except that both even and odd harmonics will be present. From the sampling frequency fs, the number of samples M, and the index of the lowest harmonic n1, estimate the fundamental frequency (first harmonic) of the vocal chords for each of your two speakers (fs*n1/M).**

Since the fundamental frequencies of different vowels are different, the vowel "u" was arbitrarily chosen to demonstrate the calculation the fundamental frequency.

For Bill, his sampling frequency fs is 20kHz, n1 is 32, and M is 2048. The fundamental frequency is therefore fs*n1/M=312.5Hz. For Dennis, the fs is also 20kHz, n1 is 29, and M is 2048. The fundamental frequency is fs*n1/M=283.2Hz. A complete table of fundamental frequencies is shown below.

Table 1 Fundamental Frequencies

| | Bill f1(Hz) | Dennis f1(Hz) |
|---|---|---|
| a | 849.6 | 664.1 |
| i | 283.2 | 293 |
| u | 312.5 | 283.2 |
| er | 449.2 | 546.9 |

**2.4 For each of the four Fourier magnitude plots (two vowels spoken by two speakers) label the first, second, and third harmonics. The kth harmonic will occur at or near Fourier index nk = k*n1. Due to the effects of windowing, each harmonic will span several adjacent Fourier coefficients.**

The four Fourier magnitude plots (for each of the two speakers) were labeled in Section 2.2.

**2.5 You should see that the magnitude of the harmonics vary with frequency. Even though all harmonics of the vocal chords are present, some frequency bands are enhanced by the position of the mouth and tongue, which form a resonant cavity. The lowest such band is the first formant and appears at frequency f1. In your four plots, identify the three most prominent formants and label their frequencies f1, f2, and f3. Compare these frequencies and the ratios f2/f1 and f3/f1 with those in Laboratory Table 22.1.**

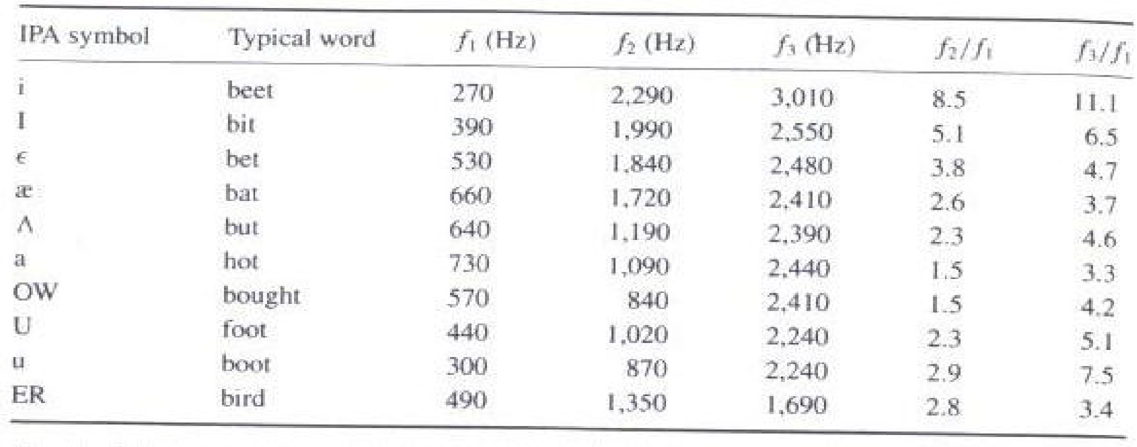
**Table 2 Measured Formant Frequencies for Bill**

| Bill | f1(Hz) | f2(Hz) | f3(Hz) |
|---|---|---|---|
| a | 849.6 | 2753.9 | N/A |
| I | 283.2 | 2334 | 3154.3 |
| U | 312.5 | 605.5 | 2627 |
| Er | 449.2 | 1337.9 | 2226.6 |

**Table 3 Measured Formant Frequencies for Dennis**

| Dennis | f1(Hz) | f2(Hz) | f3(Hz) |
|---|---|---|---|
| a | 664.1 | 2880.9 | N/A |
| i | 293 | 2304.7 | 3017.58 |
| u | 283.2 | 693.4 | 3339.9 |
| er | 546.9 | 1494.1 | 3115.2 |

Table 4 Average Formant Frequencies for Vowels [Derenzo, 464]



| IPA symbol | Typical word | $f_1$ (Hz) | $f_2$ (Hz) | $f_3$ (Hz) | $f_2/f_1$ | $f_3/f_1$ |
|---|---|---|---|---|---|---|
| i | beet | 270 | 2,290 | 3,010 | 8.5 | 11.1 |
| I | bit | 390 | 1,990 | 2,550 | 5.1 | 6.5 |
| ϵ | bet | 530 | 1,840 | 2,480 | 3.8 | 4.7 |
| æ | bat | 660 | 1,720 | 2,410 | 2.6 | 3.7 |
| Λ | but | 640 | 1,190 | 2,390 | 2.3 | 4.6 |
| a | hot | 730 | 1,090 | 2,440 | 1.5 | 3.3 |
| OW | bought | 570 | 840 | 2,410 | 1.5 | 4.2 |
| U | foot | 440 | 1,020 | 2,240 | 2.3 | 5.1 |
| u | boot | 300 | 870 | 2,240 | 2.9 | 7.5 |
| ER | bird | 490 | 1,350 | 1,690 | 2.8 | 3.4 |

Table 5 Average Formant Frequencies for Vowels Simplified

| Target | f1(Hz) | f2(Hz) | f3(Hz) |
|---|---|---|---|
| a | 730 | 1090 | 2440 |
| i | 270 | 2290 | 3010 |
| u | 300 | 870 | 2240 |
| er | 490 | 1350 | 1690 |

Table 6 Frequency Ratio Comparisons for f2/f1

|  | Target f2/f1 | Bill f2/f1 | Dennis f2/f1 | Bill Ratio Difference | Dennis Ratio Difference |
|---|---|---|---|---|---|
| a | 1.493151 | 3.241408 | 4.338051498 | 1.170851043 | 1.905300545 |
| i | 8.481481 | 8.241525 | 7.865870307 | −0.028291762 | −0.072582977 |
| u | 2.9 | 1.9376 | 2.448446328 | −0.331862069 | −0.155708163 |
| er | 2.755102 | 2.978406 | 2.731943683 | 0.081051087 | −0.008405626 |
|  |  |  | Avg Diff | 0.222937075 | 0.417150945 |

Table 7 Frequency Ratio Comparisons for f3/f1

|  | Target f3/f1 | Bill f3/f1 | Dennis f3/f1 | Bill Ratio Difference | Dennis Ratio Difference |
|---|---|---|---|---|---|
| a | 2.24 | N/A | N/A | N/A | N/A |
| i | 1.31 | 1.351457 | 1.31 | 0.028184686 | −0.00387606 |
| u | 2.57 | 4.338563 | 4.82 | 0.685066946 | 0.870771998 |
| er | 1.25 | 1.66425 | 2.09 | 0.329430429 | 0.665533346 |
|  |  |  | Avg Diff | 0.347560687 | 0.510809761 |

The average frequency ratio difference of f2/f1 is 0.22 for Bill and 0.42 for Dennis. The average frequency ratio difference of f3/f1 is 0.35 for Bill and 0.51 for Dennis.

## 3. Discussion

### 3.1 Discuss your measurement of sampling frequency (procedure section 5) and the accuracy of the measurement.

The frequency of the FPGA is exact, which means the sampling frequency is at 20 kHz. This is because FPGA has a clock on the circuit, which provides accurate timing. In older sampling system, the sampling frequency might be interrupted by high level software, which introduces delay in the measurements.

### 3.2 Examine the four Fourier coefficient data sets (procedure section 6). Discuss which features of the Fourier magnitude plot are characteristic of the vowel sound and which are characteristic of the speaker.

In the four Fourier coefficient data sets, the frequency ratios (f3/f1 and f2/f1) are characteristics of the vowel sounds. In the ideal case, the measured frequency ratio should be close to the ideal frequency ratio. The results show that lower frequency ratios (all the f2/f1) tend to be closer to the ideal ratio. The higher frequency ratios (f3/f1) ratio tends to have larger differences between the measured and ideal frequency ratio.

The absolute frequencies of the fundamental frequency and its harmonics are characteristics of the speakers. If a person has a low-pitch voice, his absolute frequencies for f1, f2, and f3 should be lower. From the measured results, Dennis tends to have higher harmonic frequencies than Bill.

### 3.3 Discuss procedure section 7. Consider sound fidelity in terms of frequency and amplitude accuracy, and higher harmonics produced by the D/A. Consider the digital storage of one hour of high-fidelity stereo music.

We did not do this part of the lab. (High fidelity audio amplifiers are supposed to be perfectly linear over a wide range of input frequencies and amplitudes)

## 4. Questions

### 4.1 To what frequency (in hertz) did the Fourier coefficient $H_{M/2}$ correspond?

$H_{M/2}$ corresponds to 10kHz, which is half of the sampling frequency. This is also called the Nyquist rate, which is the frequency at which sampled output can accurately reproduce the input signal.

### 4.2 How well did the ratio of formant frequencies f2/f1 and f3/f1 compare with Laboratory Table 22.1?

As shown in Table 6 and Table 7, the average frequency ratio difference of f2/f1 is 0.22 for Bill and 0.42 for Dennis. The average frequency ratio difference of f3/f1 is 0.35 for Bill and 0.51 for Dennis. Higher frequency ratios (f3/f1) tend to have larger differences from the ideal ratio than lower frequency ratios (f2/f1).

**4.3 For the two speakers, what were the frequency differences (in hertz) between neighboring vocal chord harmonics?**

The differences between neighboring vocal chord harmonics are shown in the tables below. The units are in Hertz.

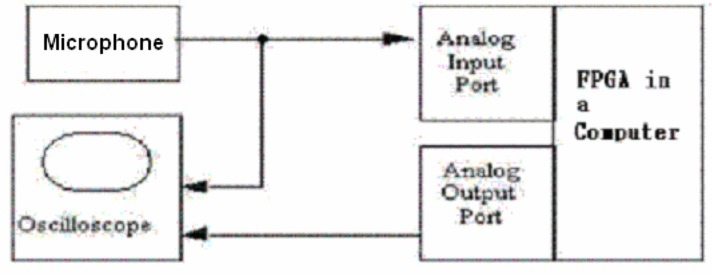
Table 8 Frequency differences between Neighboring Harmonics for Bill

| Bill | f2−f1 (Hz) | f3−f2(Hz) |
|---|---|---|
| A | 1904.3 | N/A |
| I | 2050.8 | 820.3 |
| U | 293 | 2021.5 |
| Er | 888.7 | 888.7 |

Table 9 Frequency differences between Neighboring Harmonics for Dennis

| Dennis | f2−f1(Hz) | f3−f2(Hz) |
|---|---|---|
| a | 2216.8 | N/A |
| i | 2011.7 | 712.88 |
| u | 410.2 | 2646.5 |
| er | 947.2 | 1621.1 |

**4.4 How would you design a computer program to determine which vowel was spoken, independent of speaker? Show your answer in a list of steps and comment on any major problems.**



Circuit Setup

The setup is similar to that of Lab 22 (the oscilloscope and an anti aliasing filter are optional). The program will activate the FPGA circuit to take samples exactly at 20kHz. After taking 2048 samples, the data will be transferred from the FPGA to the microcomputer through the I/O port.
The microcomputer takes the sampled data in the time domain and multiplies the sample with a Hann window. Then the time-domain data goes through a Fast Fourier Transformation (FFT) to get complex frequency values Hn. Then Hn is converted to magnitude values by

$$Fn = \text{sqrt} ( [ Re(Hn)]^2 + [Im(Hn)]^2)$$

Using an algorithm to scan the Fn values to identify peaks at n1, n2, and n3. By converting the index n1, n2, and n3 to frequency f1, f2, and f3, f2/f1 and f3/f1 ratios can be calculated. From the calculated f2/f1 and f3/f1 ratios, vowels can be identified using the ideal frequency ratio values in Table 4 Average Formant Frequencies for Vowels [Derenzo, 464]Table 4.
[Reference, Spring 2004 Midterm 2 #3d]

The major problem includes

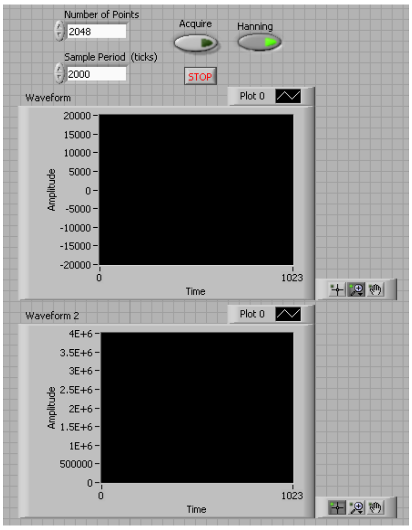
1. The difficulty to find an algorithm that realizes the highest peaks as the harmonics instead of the peaks caused by noise. For example, if the ideal fundamental frequency of "a" is 730Hz and f2/f1 of 1.49, and the speaker has an exceptional voice with a fundamental frequency of "a" at 400Hz and f2/f1 of 3.64. The program has to be designed such that it can correctly recognize the "a" from the speaker.

2. The frequency resolution with 2048 samples is 20 kHz/2048 ~= 9.77 Hz. However, the ideal frequency in Table 4 are not all integer multiples of 9.77 Hz. This introduces error when determining the signal frequency. One way to solve this problem is to sample longer, which will give a better resolution. The longer sampling period will result in more sampled data, which will require more memory to store the sampled data.

## 5. Laboratory Data Sheets

### LabView VI

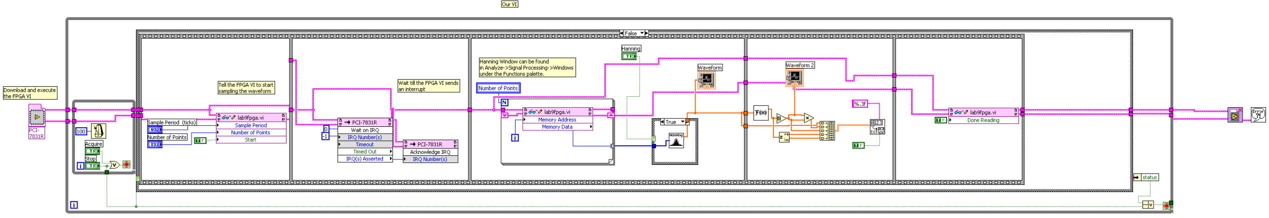
### host.vi Front Panel



Number of Points 2048

Sample Period (ticks) 2000

### Host.vi Block Diagram



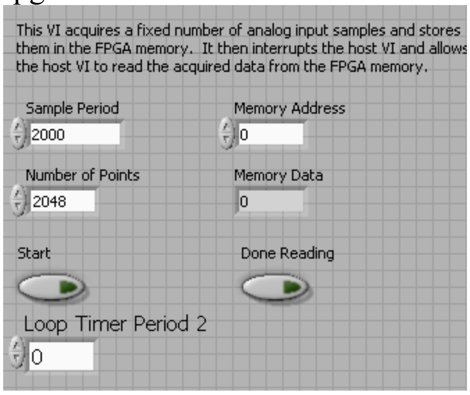
Download and execute the FPGA VI

Tell the FPGA VI to start sampling the waveform

Wait till the FPGA VI sends an interrupt

Hanning Window can be found in Analyze->Signal Processing->Windows under the Functions palette.

### fpga.vi Front Panel



This VI acquires a fixed number of analog input samples and stores them in the FPGA memory. It then interrupts the host VI and allows the host VI to read the acquired data from the FPGA memory.

Sample Period 2000

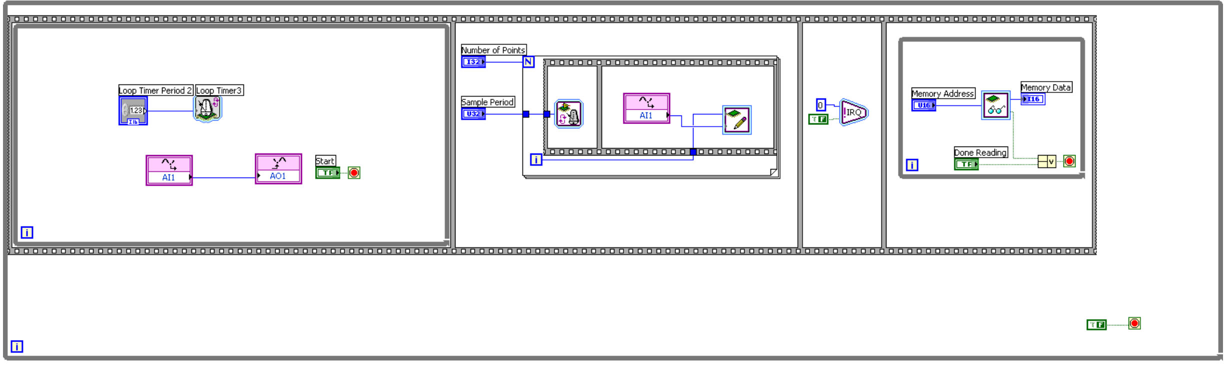
Memory Address 0

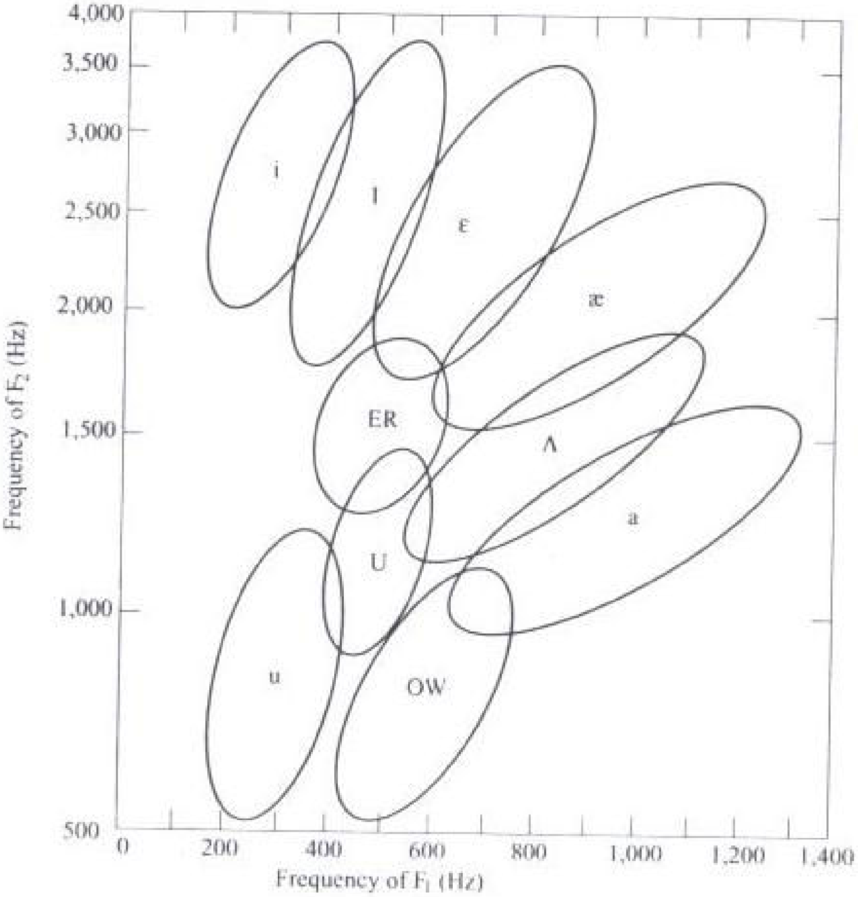
Number of Points 2048

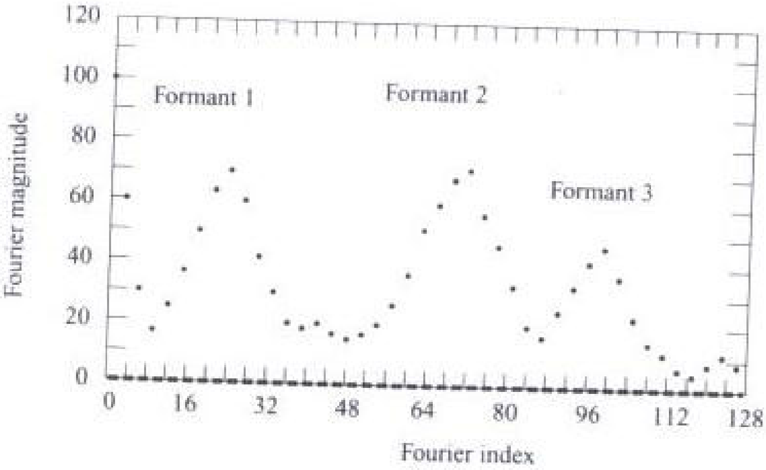
Memory Data 0

Loop Timer Period 2 0

### fpga.vi Block Diagram




4,000
3,500
3,000
2,500
2,000
1,500
1,000
500
Frequency of F₂ (Hz)
0
200
400
600
800
1,000
1,200
1,400
Frequency of F₁ (Hz)
i
ɪ
ɛ
æ
ER
ʌ
a
U
u
OW


120
100
80
60
40
20
0
Fourier magnitude
Formant 1
Formant 2
Formant 3
0
16
32
48
64
80
96
112
128
Fourier index