EE 318 Electronic Design Lab-I project report

Vowel Synthesizer

Electrical Engineering Department
IIT Bombay
April 2009

Group No. D04
Tushar Dadhich(06007027) <z6d07027@iitb.ac.in>
Amit Mangal(06D07028) <amitmangal@iitb.ac.in>
Shubham Shukla(06D07029) < sarkar.shukla@iitb.ac.in >

Faculty supervisor: Prof. S.D.Agashe
Course Instructor: Prof. Vivek Agarwal

Abstract :-

Spectrum of vowels consists of formants. Using the spectral analysis of different vowels, there formants can be determined. Thus the required envelope of any particular Vowel can be regenerated by appropriately adding frequency bands corresponding to these formants of certain bandwidth.
Here after analysis of spectrum of vowels different formants of a particular vowel are found. Then using this information, bands of the corresponding resonant frequency (formants) are produced using filters and then those bands are added to get required envelop. This envelop is then fed to speaker after amplification which sounds like a vowel.

INTRODUCTION

WHAT??

The aim of this project was to construct a device that could emulate human sounds which are called **Vowels**. Vowels are basic sounds which are produced by the vocal cord. They are unavoidable parts of human speech, without the help of which hardly any sound can be produced. **Synthesizer** is an electronic instrument capable of producing a variety of sounds by generating and combining signals of different frequencies.

WHY??

The advent of new technology has led to an increase in the number of different ways machines use to communicate with humans. Artificial voice synthesizers are highly useful in the field of human computer interaction. Text to speech synthesizers are commonly used to help people with disabilities. Generating vowels is the first step towards human voice synthesizers. Consonants are comparatively more difficult to generate than vowels.

HOW??

The spectral analysis of Vowels shows peak of certain frequencies dominant more than of the others. This type of spectrum can be created using Filters and adders on a particular type of input spectrum.
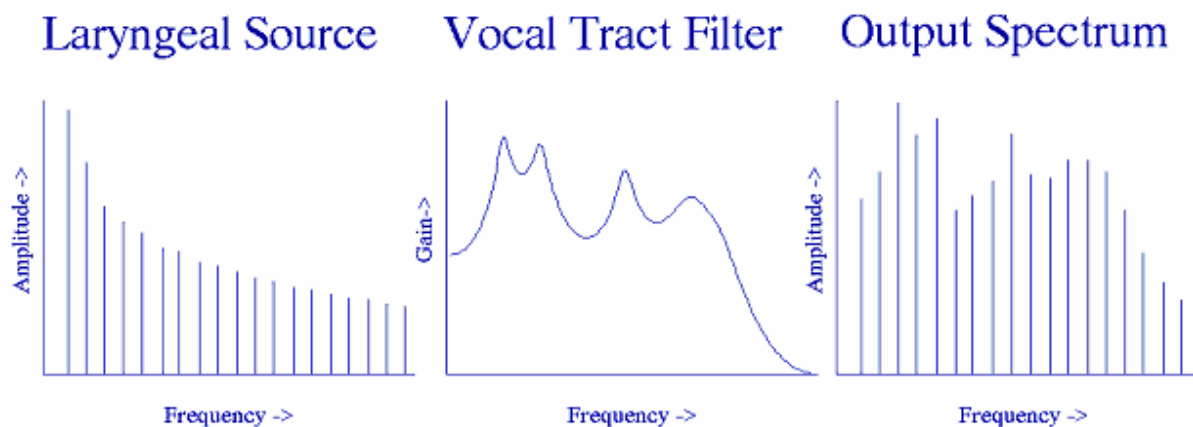
ANALYSIS

Human Voice Generation:-

When the vocal folds vibrate, they produce sound. But they vibrate the same for all voiced sounds (within a phonation type), always yielding the same "buzz". So speech, with its variety of sounds, relies on having both a **source** of sound and a means to shape or **filter** that sound.

Source: - Vocal fold vibration produces a complex periodic wave (pressure fluctuation). The spectrum of this wave contains energy at the fundamental frequency of laryngeal vibration and at multiples of the fundamental frequency harmonics.

Filter: - Energy originating at the larynx has to pass through the upper vocal tract, which is a sort of chamber or container, in order to eventually reach a listener. Resonators can function as acoustic filters. Source (input) frequencies close to a resonant frequency are amplified; other frequencies are attenuated.


The vocal tract (or actually, air in the vocal tract) has certain resonances. We call these Formants. Thus the vocal tract is a complex filter, and the formants are peaks in the vocal Tract's filter function.
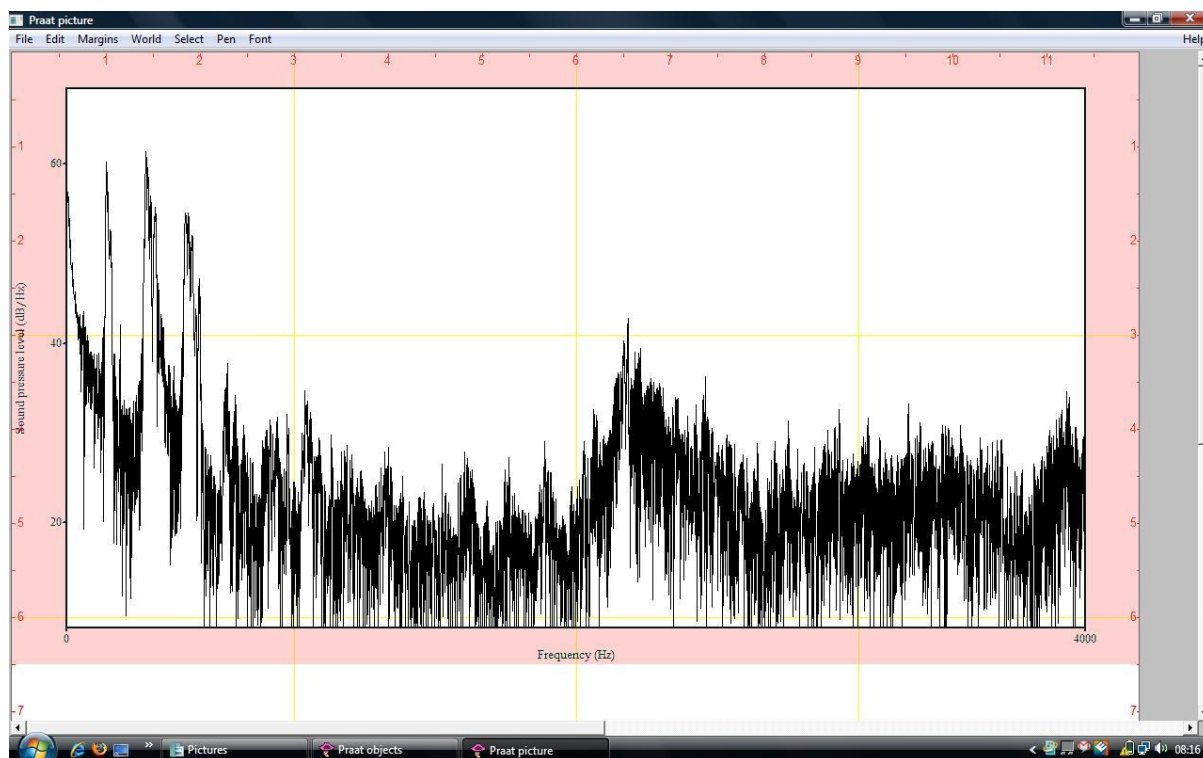
Vowel Analysis:-

The software PRAAT is designed for spectral analysis of different sounds. In our case we used it for the analysis of Vowel spectrum. Recordings were done for few samples of human voices depicting vowels.

In real life situations the formants for any of the samples of human vowels in human voice may vary substantially depending upon gendre, age etc. The results of spectral analysis of these recorded samples were done, the results varied from some ideal value, but samples showed some close relations in there spectrum i.e. the formants for the range of 2KHz to 4KHz almost occur at the same frequency .This shows that there is some correlation between the vowel spectrum produced by different people as the basic procedure of producing a particular sound is the same.
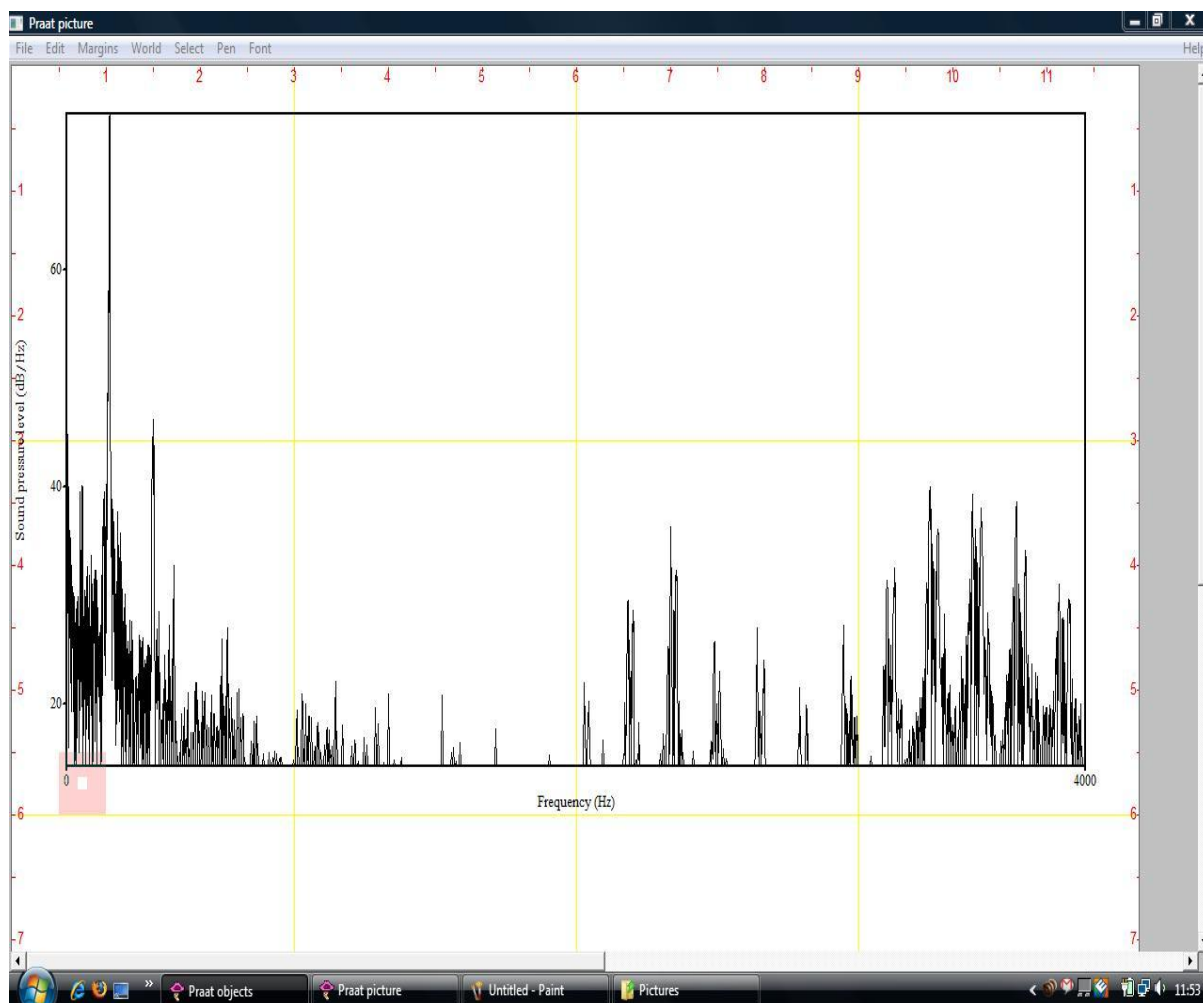
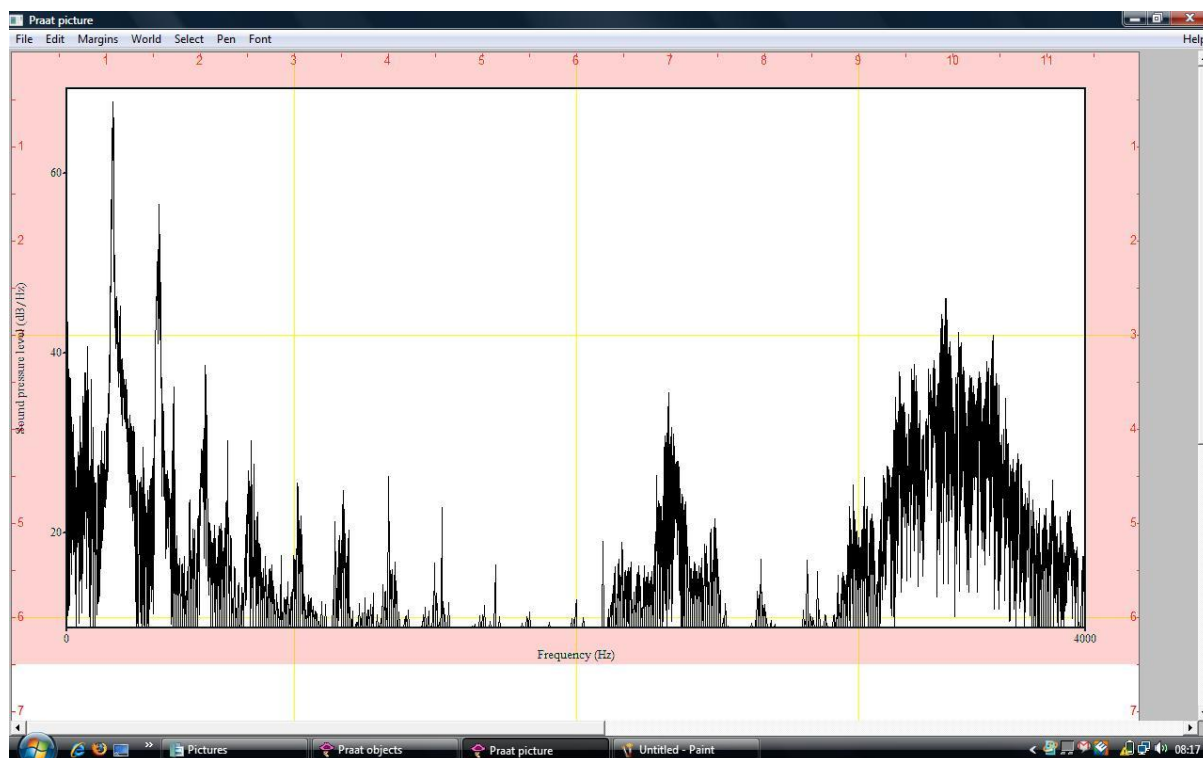The results of spectral analysis and formant analysis are attached for reference
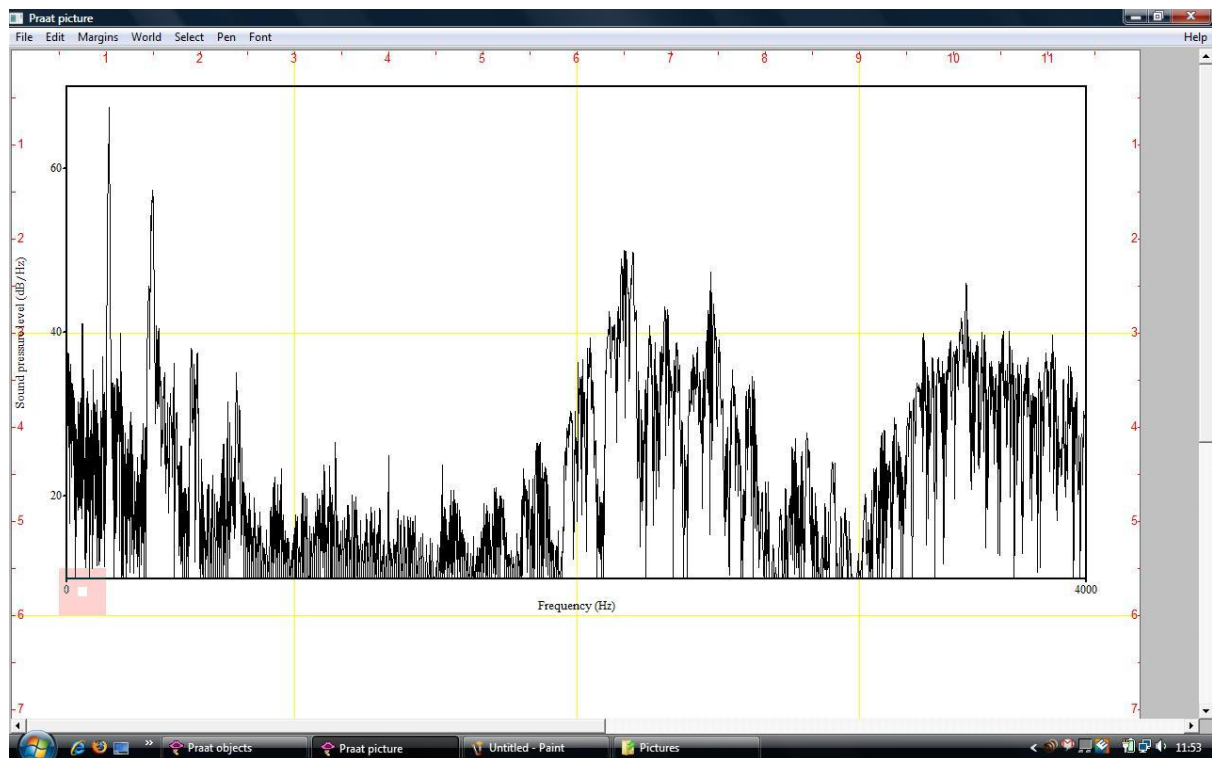
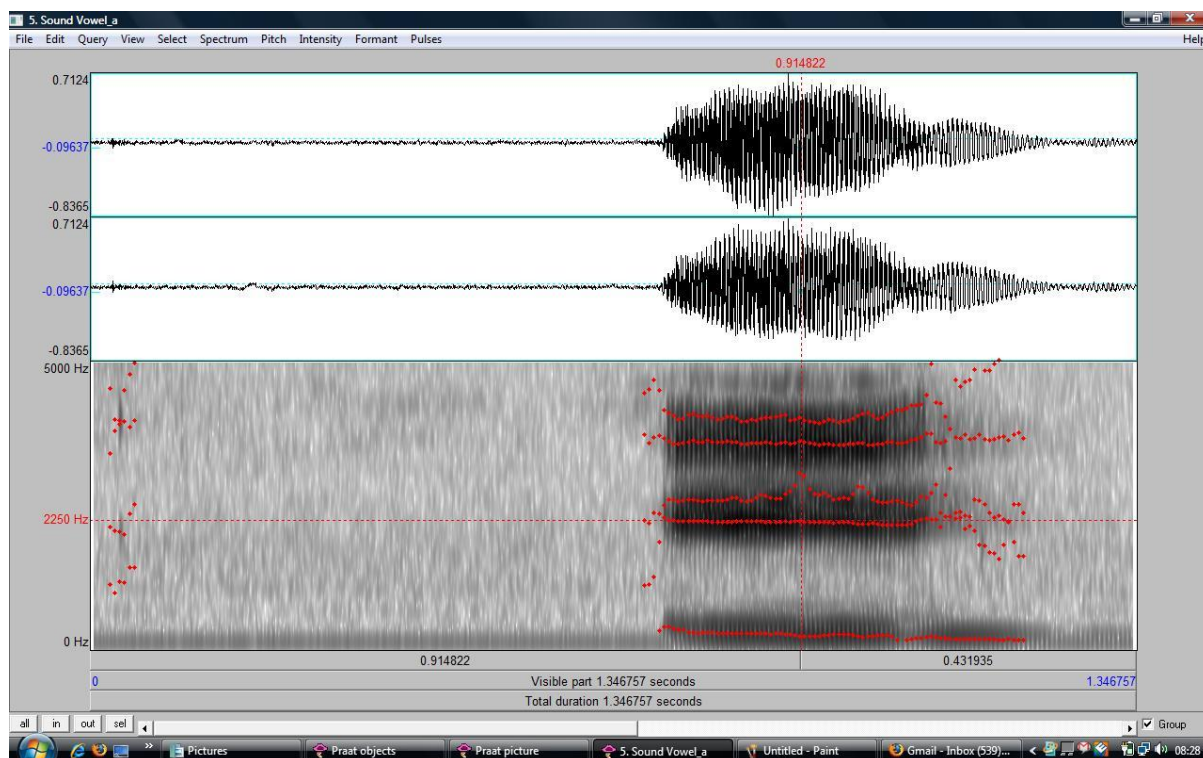Vowel "a" Spectrum for Person A
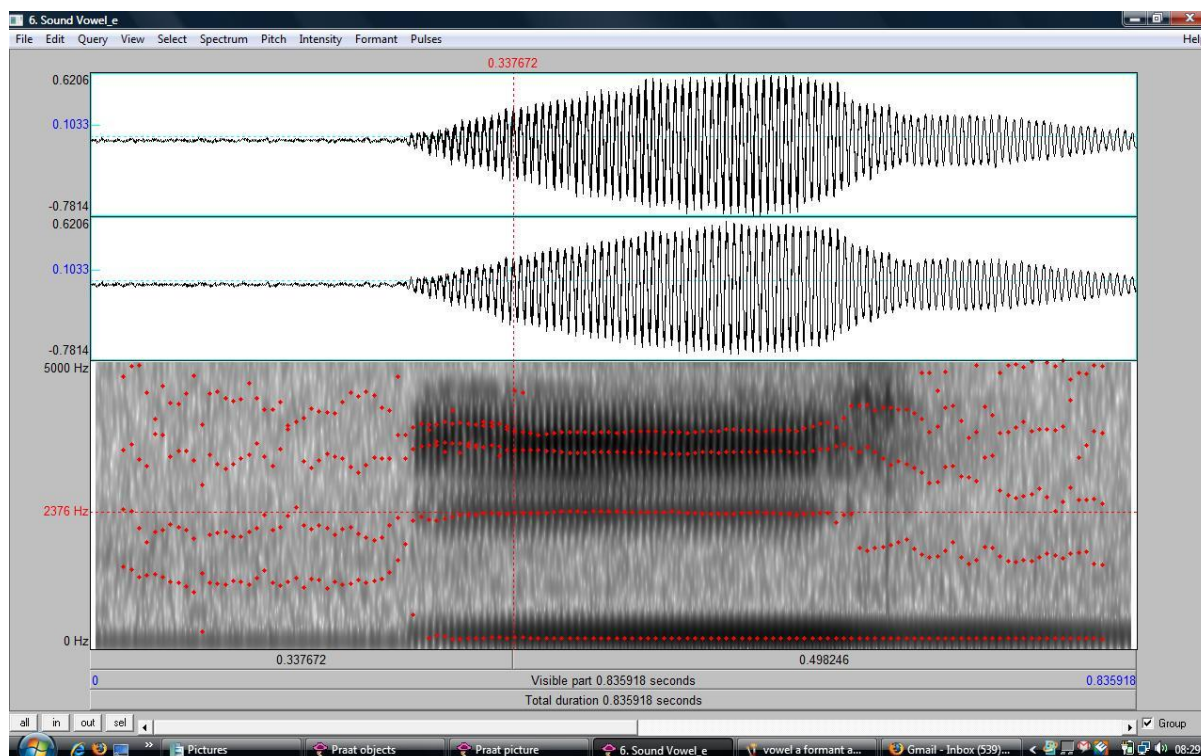
Vowel "a" Spectrum for Person B

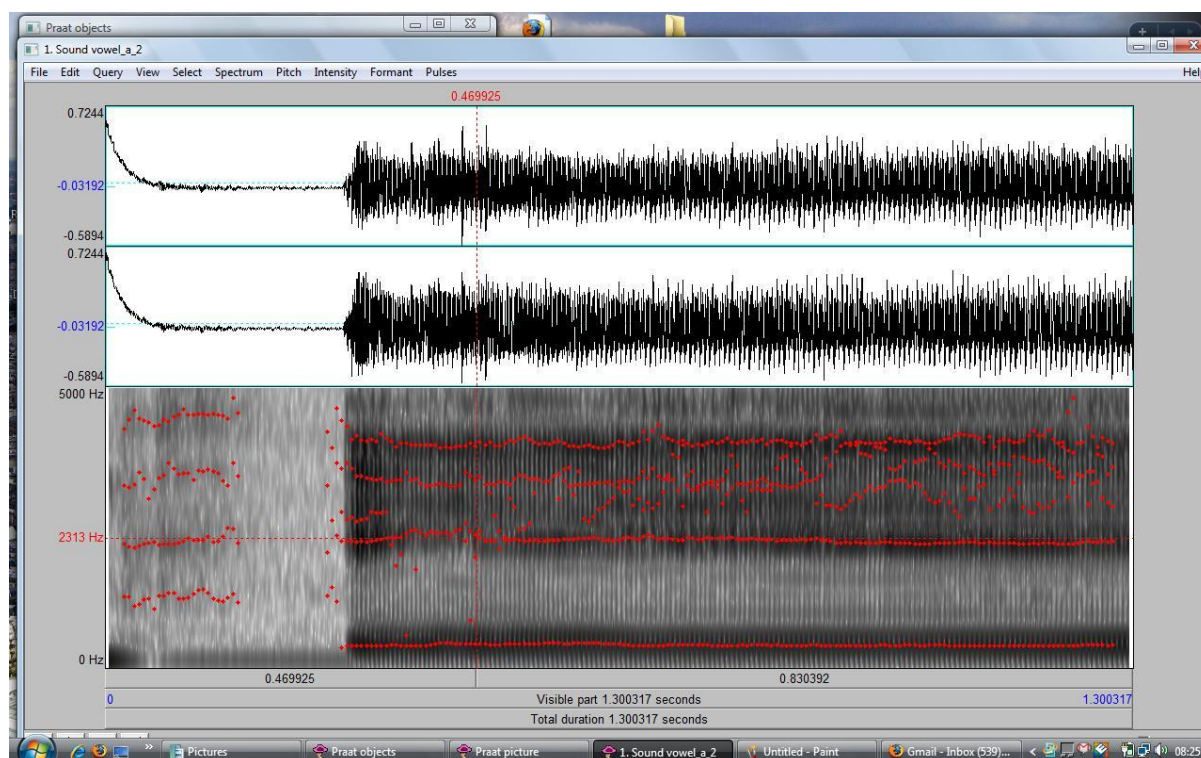Vowel "e" Spectrum for Person A

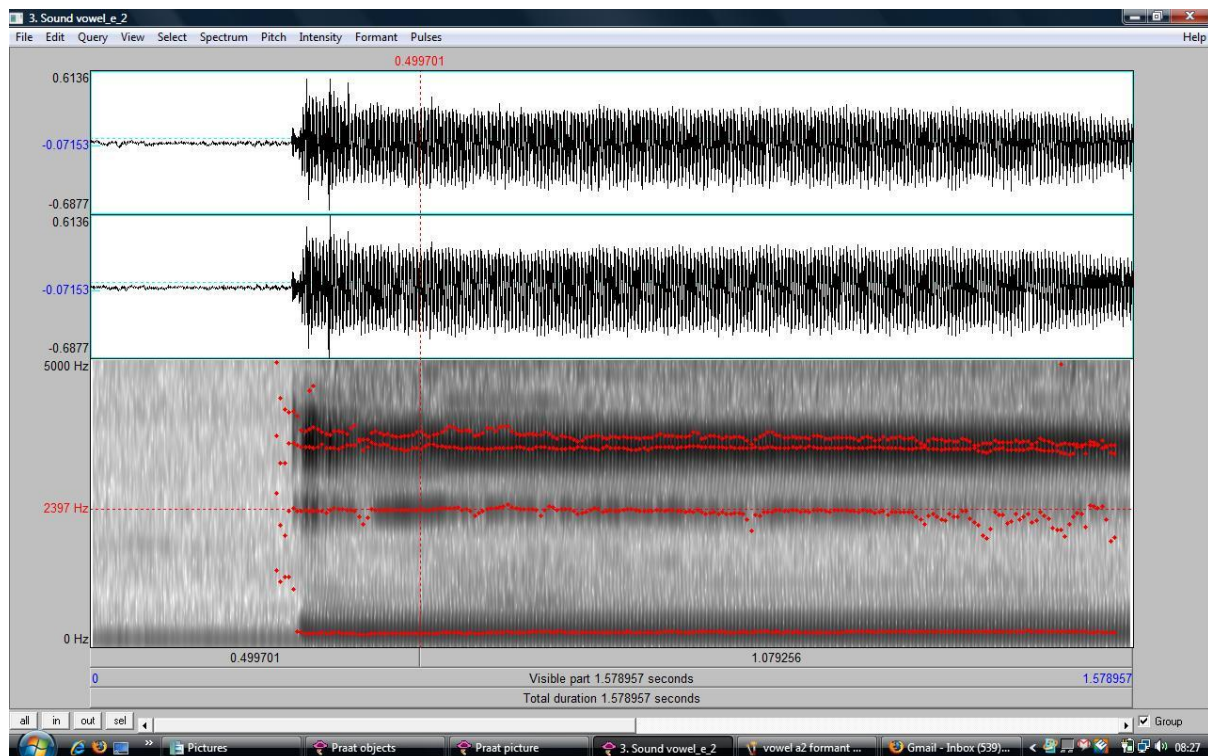Vowel "e" Spectrum for Person B

Vowel 'a' formant analysis for person A

Vowel 'e' formant analysis for person A


Vowel 'a' formant analysis for person B

Vowel 'e' formant analysis for person B

Formants:-A formant is a concentration of acoustic energy around a particular frequency in the speech wave. There are several formants, each of a different frequency, roughly one in each 1000Hz band i.e. formants occur at roughly 1 KHz intervals. Each formant corresponds to a resonance in the vocal tract.

BLOCK DIAGRAM
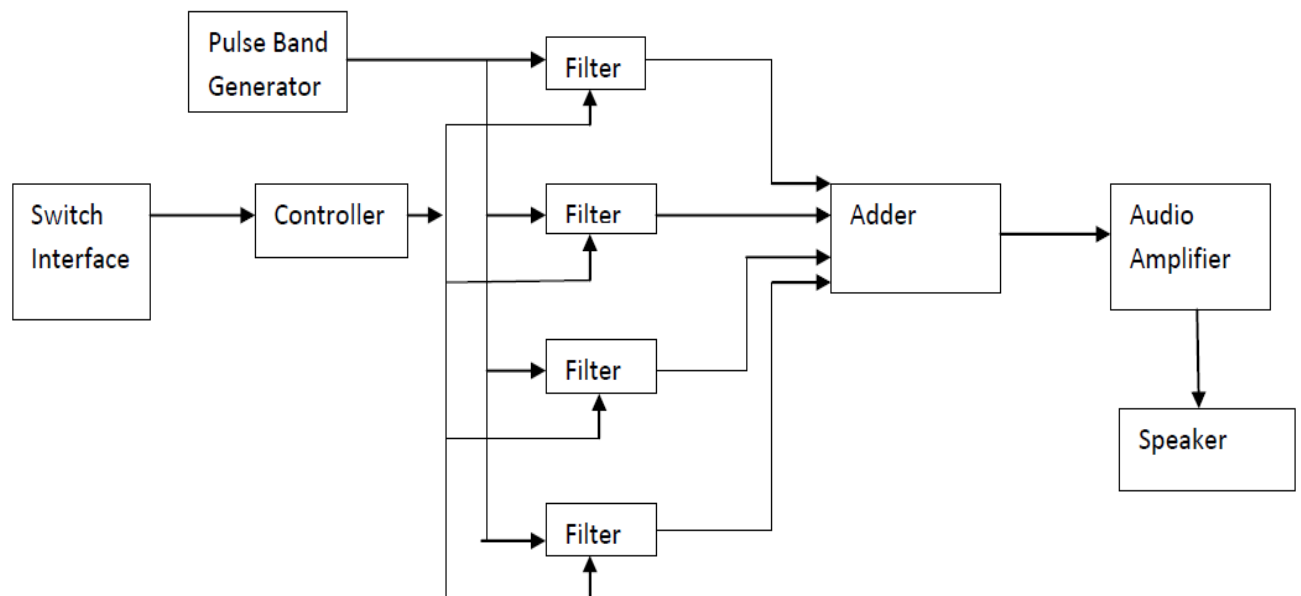
The pulse band that is to be filtered is generated by microcontroller (AT mega 16), also the clocks required to operate the filter IC's (MF-10) are generated by Microcontroller. The pulse band generated is fed into the filter IC's to produce resonant peaks which are added with help of adder to obtain before mentioned envelope. This envelope is the required output. Output is fed to audio amplifier which gives amplified signal to the speaker.

Audio Amplifier:-
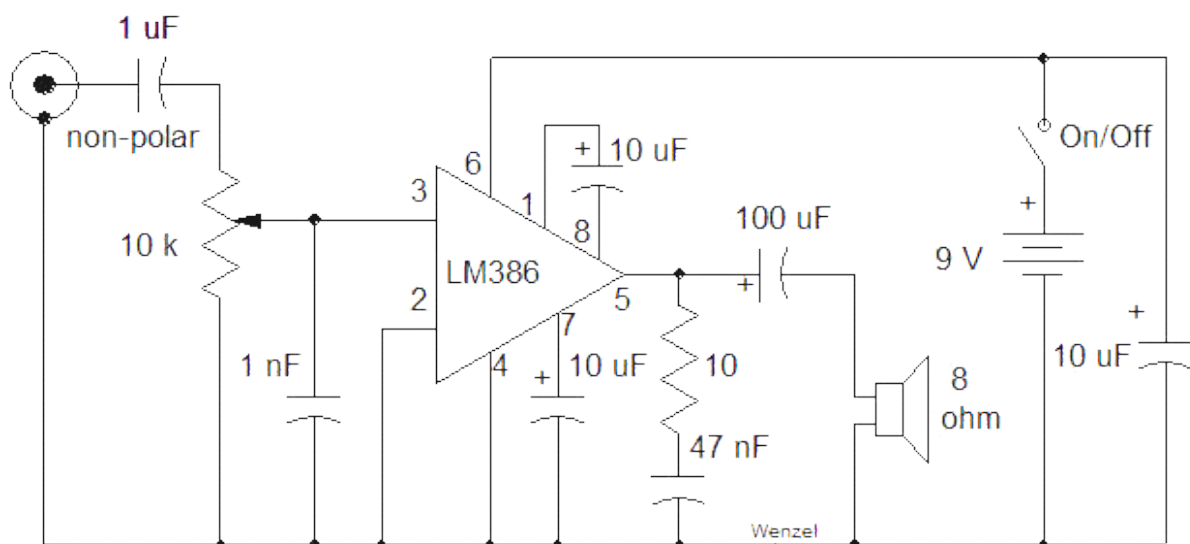
Audio amplifier using LM 386

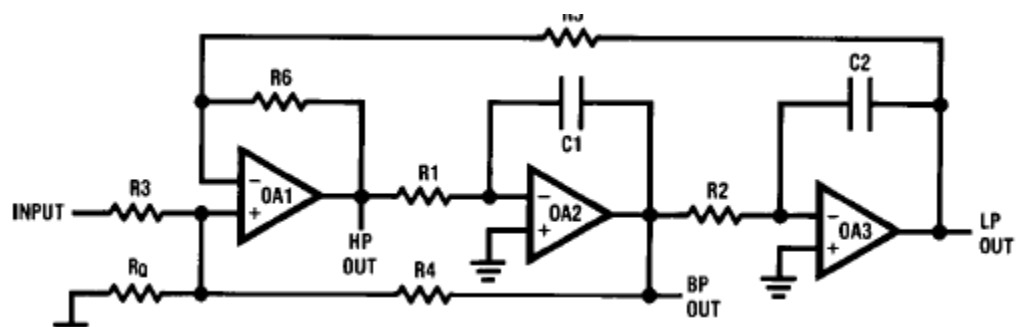<div align="center">Experimental observations:</div>

input = 20 m V peak to peak sine wave

output = 420 m V peak to peak sine wave

for all frequencies in range of 20 Hz – 2 KHz

 Shows amplification = 21 times

Filters: - Filters are needed to generate envelop. From analysis we can say that filters should be band pass filters in the frequency range of 20 Hz to 3 KHz. For this purpose we are using second order Butterworth filters (MF 10).

<div align="center">Second order filter diagram</div>

$$f_0 = \frac{1}{2\pi} \sqrt{\frac{R6}{R5}} \sqrt{\frac{1}{R1\,R2\,C1\,C2}}$$

$$Q = \left(\frac{1 + \dfrac{R4}{R3} + \dfrac{R4}{R0}}{1 + \dfrac{R6}{R5}} \right) \sqrt{\frac{R6\,R1\,C1}{R5\,R2\,C2}}$$
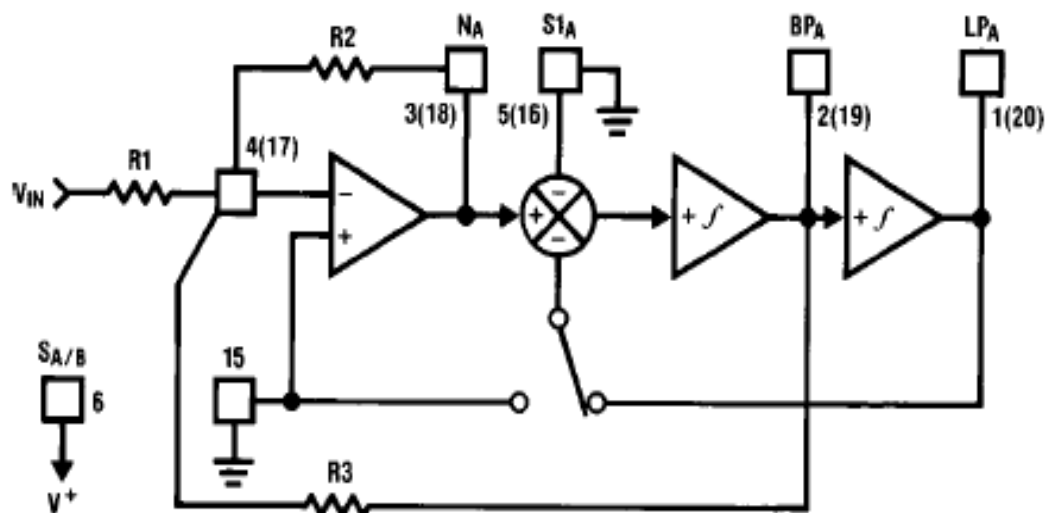
Design used for filters

Design Equations

$$f_o = \frac{f_{CLK}}{100} \text{ or } \frac{f_{CLK}}{50}$$

$$f_{notch} = f_o$$

$$Q = \frac{R3}{R2}$$

$$H_{OBP} = -\frac{R3}{R1}$$

Here values of resistors used are R1=3.3K, R2=1K, R3=10K

Experimental observations:
For sine waves
f_c = 5 KHz
f_o = 100 Hz

Freq (in) (Hz)	amplitude (in) (Volts)	freq (out) (Hz)	amplitude (out) (Volts)
80	0-5	80	-2 to 2
90	0-5	90	-3 to 3
100	0-5	100	-4 to 4
110	0-5	110	-3 to 3
120	0-5	120	-2 to 2
140	0-5	140	-1 to 1
160	0-5	160	- 0.7 to 0.7

f_c = 25 KHz
f_o = 500 Hz

Freq (in) (Hz)	amplitude (in) (Volts)	freq (out) (Hz)	amplitude (out) (Volts)
400	0-5	400	-1.8 to 1.8
450	0-5	450	-3.2 to 3.2
500	0-5	500	-4.2 to 4.2
550	0-5	550	-3.4 to 3.4

f_c = 100 KHz
f_o = 2 KHz

| Freq (in) (KHz)	amplitude (in) (Volts)	freq (out) (KHz)	amplitude (out) (Volts)
1.5	0-5	1.5	-1.2 to 1.2
1.8	0-5	1.8	-3.2 to 3.2
1.9	0-5	1.9	-4.0 to 4.0
2.0	0-5	2.0	-4.0 to 4.0
2.1	0-5	2.1	-4.0 to 4.0
2.2	0-5	2.2	-3.4 to 3.4

Microcontroller:

The microcontroller used is ATmega16. It was initially supposed to generate all the four formant frequencies (actually 50 times of required) serving as clocks to filters in addition of providing input for them. A code was written for the task which essentially was polling the pins of port D looking for any pin being high. The pins of port D were connected to the switches. As soon as a pin is found high the corresponding frequencies were to be generated at port B pins by comparing the counter value to some fixed values and then toggling the pins of port B. The code had some issues and it did not run.

Therefore, another totally different code was thought which had all the timers of MuC in use. In this code one of the timers operate in fast PWM mode giving the pulse band which is input to the filters. Other two timers were used in CTC mode giving the required clocks for filters. Again continues polling of port D pins was done in order to find the key pressed and accordingly initialize the timers and get required output. It was found that the code ran ok for checking of just one pin but not for the polling process. Also it was found that the timer providing pulse band gave the required output irrespective of which code was used. Hence the code had to be discarded.

After some more alterations in code it was thought that we should use the external interrupt facility in the MuC. A code was written that decided the interrupt status and a counter variable that rose from 0 to 5 and again to 0.The cases thus got gave timer initializations and other counter variable initialization that gave rise to another frequency on some pin by toggling. Unfortunately this code didn't gave any results compelling to use the previous meathod of three timers that too for just one vowel.

Adder: - NON inverting Adder is used for the required purpose.
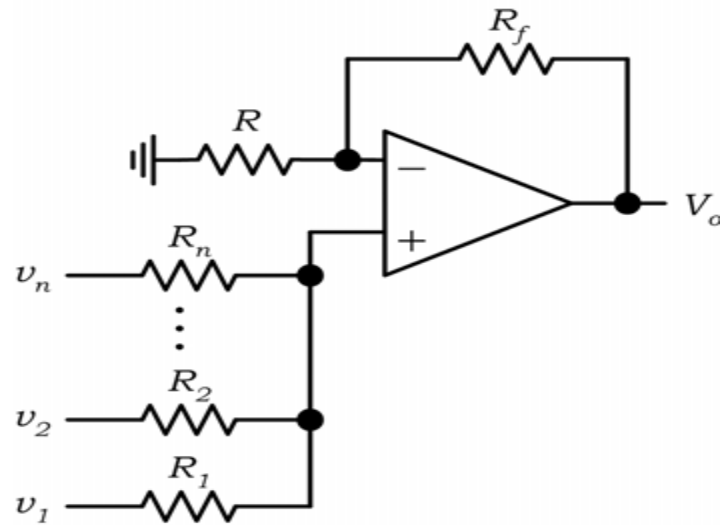
$$V_o = +K \sum_{i=1}^{n} A_i v_i$$

$$\text{where, } K = \left(1 + \frac{R_f}{R}\right) \frac{R_p}{R_f},$$

$$A_i = \frac{R_f}{R_i},$$

$$R_p = R_1 \, // \, R_2 \, // \, ... \, // \, R_n$$

Here OP-07 is used as an op-amp IC, because it has very high input resistance (140 G ohms). High input resistance was required as filter IC's were unable to provide required current.

CONCLUSION

Though some sound depicting to that of vowels were produced but they were not very much similar to that of the human voice, mainly due to the fact that pulse band used in the project is not similar to that of glottal waves produced. Also producing resonant bands and adding them is not a good way as can be seen that real vowel spectrum consists of more peaks then only formants.
So although this method can produce sound close to that of humans, but it is not a good method.

REFERRENCES

Books by Authors:
        Flanagan
        O' shaughnessy, Douglas

Web references:
        http://www.stanford.edu/class/linguist205/index_file
        www.**datasheet**catalog.com/**datasheets**_pdf
        ftp://trf.education.gouv.fr/pub/edutel/siac/siac2/jury/2004/