NEAR EAST UNIVERSITY

FACULTY OF ENGINEERING

DEPARTMENT OF ELECTRICAL AND ELECTRONIC ENGINEERING

SPEECH SIGNAL PROCESSING

GRADUATION PROJECT
EE-400

STUDENT: MOHAMMAD AL-DERHALLI (20010989)

SUPERVISOR: Prof.Dr. Fakhreddin Mamedov

Nicosia-2002

TAPLE OF CONTENTES

ACKNOWLEDGEMENT

Firstly I would like to present my special appreciation to my supervisor Prof.Dr. Fakhreddin Mamedov, without whom it is not possible for me to complete the project. His trust in my work and me and his priceless awareness for the project has made me do my work with full interest. His friendly behavior with me and his words of encouragement kept me doing my project.

Secondly I offer special thanks to my parents, who encouraged me in every field of life and try to help whenever I needed. They enhanced my confidence in myself to make me able to face every difficulty easily. I am also grateful to my mother whose prayers and my father whose words for me had made this day comes true. And because of them I am able to complete my work

I would also like to pay my special thanks to my all friends who helped me and encouraged me for doing my work. Their reluctance and friendly environment for me has helped me allot. I want to thank them as they contributed their time and provided very helpful suggestions to me.

Abstract

Speech production models, coding methods as well as text to speech technology often lead to the introduction of modulation models to represent speech signals with primary components which are amplitude-and-phase-modulated sine functions. Parallelisms between properties of the wavelet transform of primary components and algorithmic representations of speech signals derived from auditory nerve models like the EIH lead to the introduction of synchrosqueezing measures. On the other hand, in automatic speech (and speaker) recognition, cepstral feature have imposed themselves quasi-universally as acoustic

Also The success of speech recognition depends on the size of the vocabulary and the quality of the speech signal. The zero-error recognition of unrestricted, continuous speech from a noisy environment (the "electronic secretary"), however, is still in the future.

Speaker identification has helped solve several disasters (mid-air collision over the Grand Canyon, burning-up of three astronauts), but its forensic applications are limited if the pool of potential speakers is large. Speaker *verification* is of increasing importance in limiting access to restricted data (financial, medical, military).

INTRODUCTION

The aim of this project is to implement some of the digital signal processing techniques on the speech signals by using analysis and electronic workbench (EWB), which makes it easy to do.

The speech signal processing is one of the most important fields in telecommunications. There are many areas that speech processing concerns with such as digitals transmission and storage of speech, speaker verification and identification systems, speech recognition systems and speech synthesis systems.

The project can be divided into three different chapters: the first chapter isintroduction to speech processing, that is include, the basic properties of speecSignal processing and digital speech processing, such as a representation of speech signal that is into part one is a waveform representation and the other is parametric representation. In the second chapter we studied and analysis the waveform representation that it included a pulse code modulation (PCM) involves that a uniform (PCM) then signal to quantize noise ratio (SQNR) and nonuniform pulse code modulation.

The final chapter is about analysis electronics synthesis of speech by (EWB). That is including formant synthesis and analogue circuit formant synthesis. We studied in circuit formant synthesis the design of synthesis speech signal model; this chapter is in general a spectrum synthesis, that modeling speech by modeling its spectrum.

1. SPEECH SIGNAL PROCESSING

1.1 The Basic Properties of Speech:

Speech is produced when air is forced from the lungs through the vocal cords and along the vocal tract. The vocal tract extends from the opening in the vocal cords (called the glottis) to the mouth, and in an average man is about 17 cm long.

It introduces short-term correlations (of the order of 1 ms) into the speech signal, and can be thought of as a filter with broad resonances called formants. The frequencies of these formants are controlled by varying the shape of the tract, for example by moving the position of the tongue.

An important part of many speech codecs is the modeling of the vocal tract as a short term filter. As the shape of the vocal tract varies relatively slowly, the transfer function of its modeling filter needs to be updated only relatively infrequently (typically every 20 ms or so).

The vocal tract filter is excited by air forced into it through the vocal cords. Speech sounds can be broken into three classes depending on their mode of excitation.

- Voiced sounds are produced when the vocal cords vibrate open and closed, thus interrupting the flow of air from the lungs to the vocal tract and producing quasi-periodic pulses of air as the excitation.
- The rate of the opening and closing gives the pitch of the sound. This can be adjusted by varying the shape of, and the tension in, the vocal cords, and the pressure of the air behind them. Voiced sounds show a high degree of *periodicity at the pitch period, which is typically* between 2 and 20 ms. This long-term periodicity can be seen in Figure 1 which shows a segment of voiced speech sampled at 8 kHz. Here the pitch period is about 8 ms or 64 samples. The power spectral density for this segment is shown in Figure 2.
- *Unvoiced sounds result when the excitation is a noise-like turbulence* produced by forcing air at high velocities through a constriction in the vocal tract while

the glottis is held open. Such sounds show little long-term periodicity as can be seen from Figures 3 and 4 , although short-term correlations due to the vocal tract are still present.

- Plosive sounds result when a complete closure is made in the vocal tract, and air pressure is built up behind this closure and released suddenly.

Some sounds cannot be considered to fall into any one of the three classes above, but are a mixture. For example voiced fricatives result when both vocal cord vibration and a constriction in the vocal tract are present.

Although there are many possible speech sounds which can be produced, the shape of the vocal tract and its mode of excitation change relatively slowly, and so speech can be considered to be quasi-stationary over short periods of time (of the order of 20 ms).

We can see from Figures 1, 2, 3, and 4 that speech signals show a high degree of predictability, due sometimes to the quasi-periodic vibrations of the vocal cords and also due to the resonances of the vocal tract. Speech coders attempt to exploit this predictability in order to reduce the data rate necessary for good quality voice transmission.

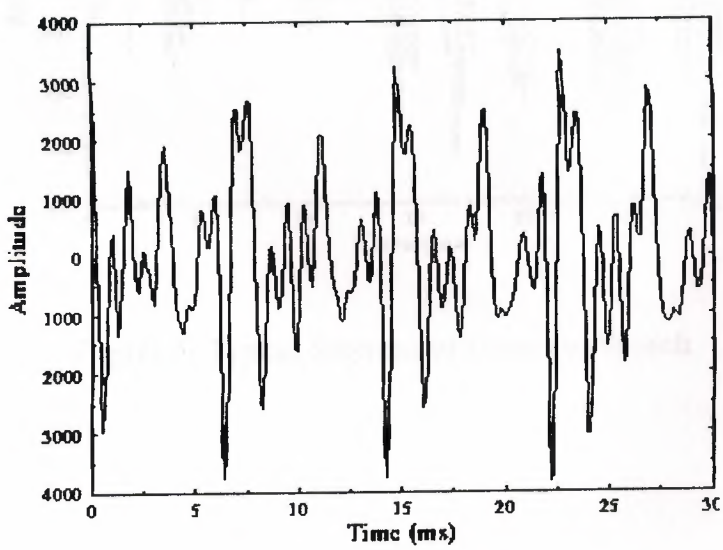

Figure 1: Typical Segment of Voiced Speech

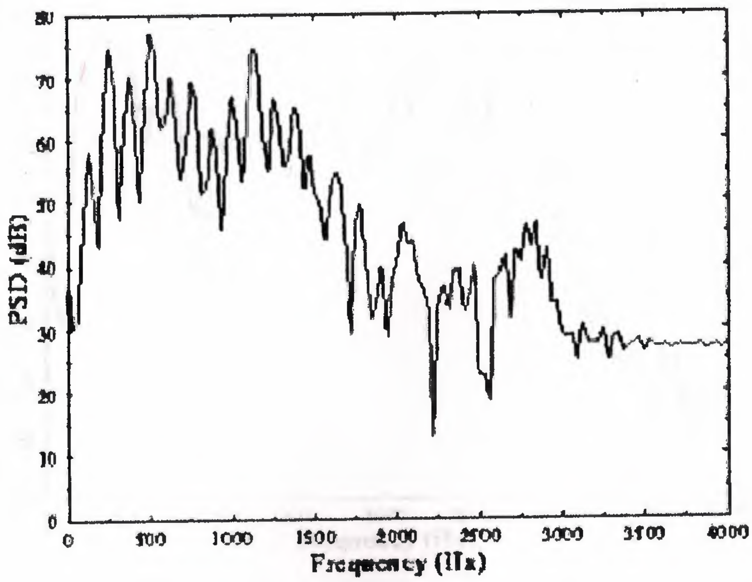

Figure 2: Power Spectral Density for a Segment of Voiced Speech

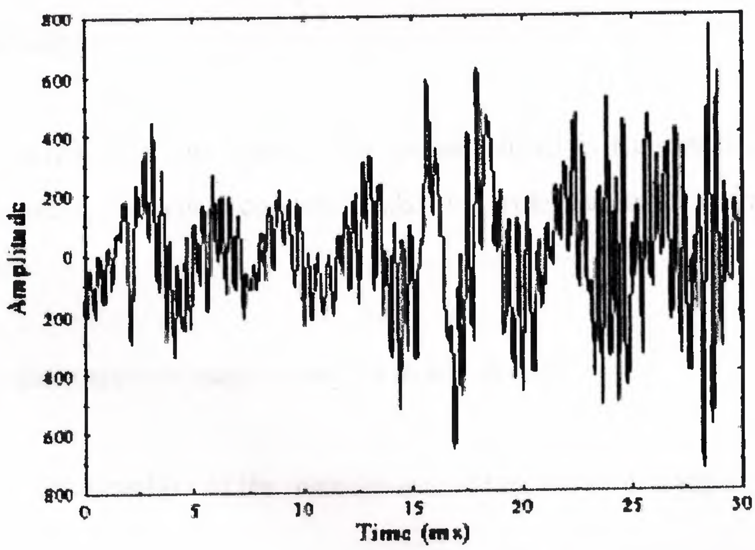

Figure 3: Typical Segment of Unvoiced Speech

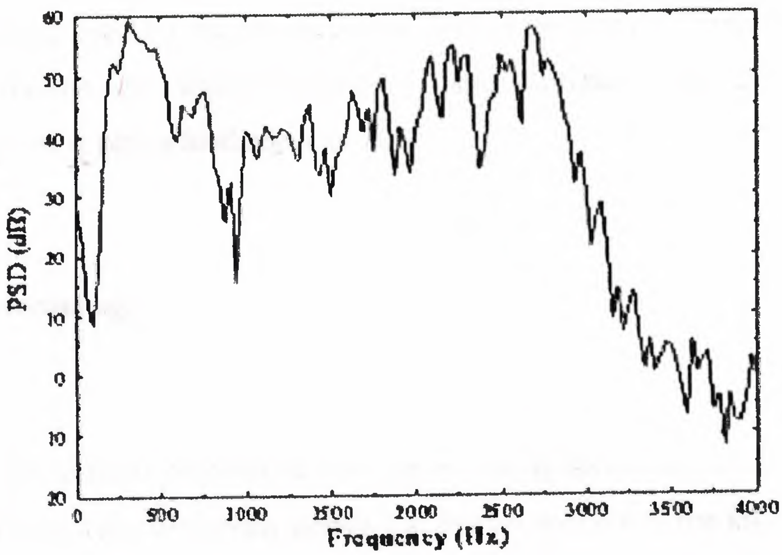

Figure 4: Power Spectral Density for a Segment of Voiced Speech

1.2 Speech processing:

In speech communication system, the speech signal is transmitted, stored, and processed in any ways. Technical concern leads to a wide variety of representations of the speech signal.

In general, there are two major concerns in any system:

1. Preservation of the massage content in the speech signal.

2. Preservation of the speech signals in a form that is convenient for transmission or storage, or in a form that is flexible so that modifications may be made to the speech signal with out seriously degrading the massage content.

The representation of speech signal must be such that the information content can easily be extracted by human listeners, or automatically by machine. Throughout this project we shall see that representation of speech signal (rather than massage content) may require from 500 to upward of 1 million bits per second. In the design and implementation of these representations, the methods of signal processing play a fundamental role.

1.3 Signal processing:

The general problem of information manipulation and processing is depicted in fig. In the case of speech signals the human speaker is the information source. The measurement or observation is generally he acoustic waveform.

Signal processing involves fit obtaining a representation of signal based on a given model and then the application of some higher-level transformation in order to put the signal into a more convenient form. The last step in the process is the extraction and utilization of the massage information. Machines may perform this step either by human listeners or automatically. By way of example, a system whose function is to automatically identify a speaker from a given set of speakers might use a time-dependant spectral representation of the speech signal. One possible signal transformation would be to average spectral across an entire sentence, compare the average spectrum to a stored average spectrum template for each possible speaker, and then based on a spectral similarly measurement chooses the identity of the speaker. For this example the "information" in the signal is identity of the speaker.

Thus, processing of speech signals generally involves two tasks. First, it is a vehicle for obtaining a general representation of a speech signal in either waveform or parametric form. Second, signal processing serves the function of aiding in the process of transforming the representation into alternate forms, which are less general in nature, but more appropriate to specific application. Throughout this project we ill see numerous specific examples of the importance of signal processing in the area of speech communication.

1.4 Digital speech processing:

In considering the application of digital signal processing techniques to speech communication problems.

It is helpful to focus on three main topics:

1. The representation of speech signal in digital form.
2. The implementation of the sophisticated processing techniques.
3. The classes of application, which rely heavily on digital processing.

The representation of speech signals in digital form is, of course, of fundamental concern. In this regard we are guided by the well-known sampling theorem which states that a bandlimited signal can be represented by samples taken periodically in time - provided that the samples are taken at a high enough rates.

Thus, the process of sampling underlies all of the theory and application of signal speech processing. There are many possibilities for discreet representation of speech signals. As shown in fig.1-3-1, these representations can be classified into two broad groups, namely waveform representation and parametric representations.

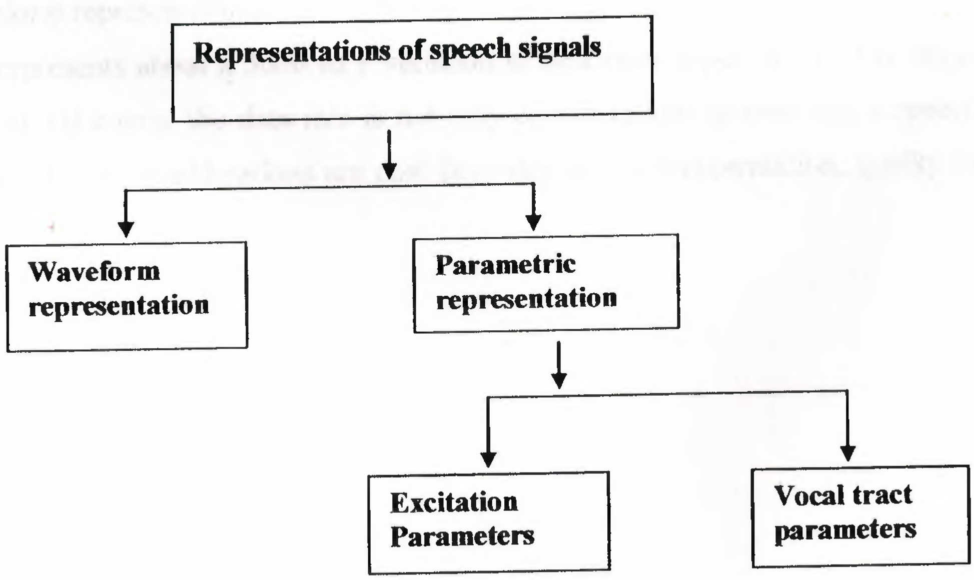

Fig.1-3-1 representation of speech signals

Waveform representation, as the name implies, are concerned with simply preserving the "wave shape" of the analogue speech signal through a sampling and quantization process.

Parametric representations, on the other hand, are concerned with representing the speech signal as the out put of a model for speech production. The first step in obtaining a parametric representation is often a digital waveform representation; that is, the speech signal is sampled and quantized and then further processed to obtain the parameters of the model for speech production.

The parameters of this model are conveniently classified as either excitation parameters (i.e., related to the Source of speech sounds) or vocal tract response parameters (i.e., related to the individual speech sounds).

Fig.1-3-2 shows a comparison of a number of different representations of speech signal according to the data rate required. The dotted line, at a data rate a bout 15,000 bits per second, separates the high data rate waveform representation at the left from the lower data rate parametric representations at the right.

This figure shows variations in data rate from 75 bits per second (approximately the basic message information of the text) to rate upward of 200,000 bits per second for simple waveform representations.

This represents about a 3000 to 1 variation in data rates depending on the signal representation. Of course the data rate is not only considerations in choosing a speech representation. Other considerations are cost flexibility of the representation, quality of speech.

200,000 60,000 20,000 10,000 500

75

LDM,PCM,DPCM,ADM analysis synthesis

synthesis

Methods from

printed text

(no Source coding) (Source coding)

Waveform representation parametric

representation

Fig.1-3-2 range of bit rates for various types of speech representations.

2. Waveform representation

Introduction:

Here we discuss the main speech coding techniques which are used today, and those which may be used in the future. In order to simplify the description of speech codecs they are often broadly divided into three classes - waveform codecs, source codecs and hybrid codecs.

Typically waveform codecs are used at high bit rates, and give very good quality speech. Source codecs operate at very low bit rates, but tend to produce speech which sounds synthetic. Hybrid codecs use techniques from both source and waveform coding, and give good quality speech an intermediate bit rates.

This is shown in Figure 5, which shows how the speech quality of the three main classes of speech codecs vary with the bit rate of the codec.

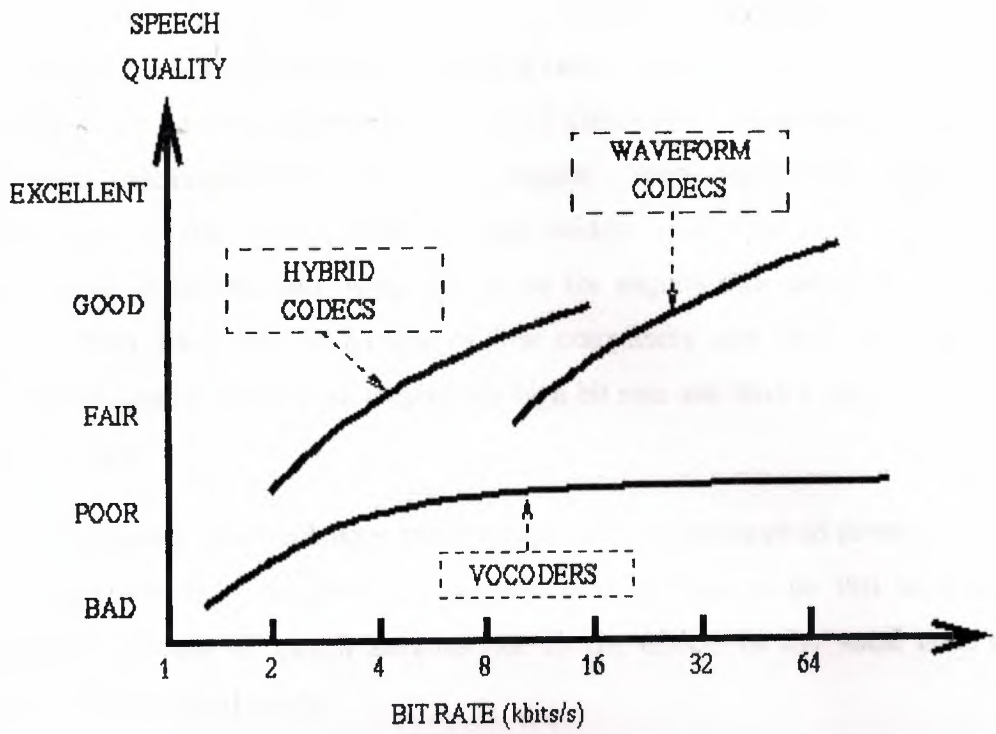

Figure 5: Speech Quality versus Bit Rate for Common Classes of Codes

2.1Waveform coding:

Waveform codecs attempt, without using any knowledge of how the signal to be coded was generated, to produce a reconstructed signal whose waveform is as close as possible to the original. This means that in theory they should be signal independent and work well with non-speech signals. Generally they are low complexity codecs which produce high quality speech at rates above about 16 kbits/s. When the data rate is lowered below this level the reconstructed speech quality that can be obtained degrades rapidly.

The simplest form of waveform coding is Pulse Code Modulation (PCM), which merely involves sampling and quantizing the input waveform. Narrow-band speech is typically band-limited to 4 kHz and sampled at 8 kHz. If linear quantization is used then to give good quality speech around twelve bits per sample are needed, giving a bit rate of 96 kbits/s. This bit rate can be reduced by using non-uniform quantization of the samples.

In speech coding an approximation to a logarithmic quantizer is often used. Such quantizer give a signal to noise ratio which is almost constant over a wide range of input levels, and at a rate of eight bits/sample (or 64 kbits/s) give a reconstructed signal which is almost indistinguishable from the original. Such logarithmic quantizer were standardized in the 1960's, and are still widely used today. In America u-law commanding is the standard, while in Europe the slightly different A-law compression is used. They have the advantages of low complexity and delay with high quality reproduced speech, but require a relatively high bit rate and have a high susceptibility to channel errors.

A commonly used technique in speech coding is to attempt to predict the value of the next sample from the previous samples. It is possible to do this because of the correlations present in speech samples due to the effects of the vocal tract and the vibrations of the vocal cords.

If the predictions are effective then the error signal between the predicted samples and the actual speech samples will have a lower variance than the original speech samples. Therefore we should be able to quantize this error signal with fewer bits than the original speech signal. This is the basis of Differential Pulse Code Modulation (DPCM) schemes - they quantize the *difference* between the original and predicted signals.

The results from such codecs can be improved if the predictor and quantizer are made adaptive so that they change to match the characteristics of the speech being coded. This leads to Adaptive Differential PCM (ADPCM) codecs. In the mid 1980's the CCITT standardized a ADPCM codec operating at 32 kbits/s, which gave speech quality that was very similar to the 64 kbits/s PCM codecs. Later ADPCM codecs operating at 16,24 and 40 kbits/s were also standardized.

The waveform codecs described above all code speech with an entirely time domain approach. Frequency domain approaches are also possible, and have certain advantages. For example in Sub-Band Coding (SBC) the input speech is split into a number of frequency bands, or sub-bands, and each is coded independently using for example an ADPCM like coder. At the receiver the sub-band signals are decoded and recombined to give the reconstructed speech signal.

The model parameters can be determined by the encoder in a number of different ways, using either time or frequency domain techniques. Also the information can be coded for transmission in various different ways. Voiceovers tend to operate at around 2.4 kbits/s or below, and produce speech which although intelligible is far from natural sounding. Increasing the bit rate much beyond 2.4 kbits/s is not worthwhile because of the inbuilt limitation in the coder's performance due to the simplified model of speech production used.

The main use of voiceovers has been in military applications where natural sounding speech is not as important as a very low bit rate to allow heavy protection and encryption

2.3 Hybrid Codecs :

Hybrid codecs attempt to fill the gap between waveform and source codecs. As described above waveform coders are capable of providing good quality speech at bit rates down to about 16 kbits/s, but are of limited use at rates below this.

Voiceovers on the other hand can provide intelligible speech at 2.4 kbits/s and below, but cannot provide natural sounding speech at any bit rate. Although other forms of hybrid codecs exist, the most successful and commonly used are time domain Analysis-by-Synthesis (AbS) codecs.

Such coders use the same linear prediction filter model of the vocal tract as found in LPC voiceovers. However instead of applying a simple two-state, voiced/unvoiced, model to find the necessary input to this filter, the excitation signal is chosen by attempting to match the reconstructed speech waveform as closely as possible to the original speech waveform.

AbS codecs were first introduced in 1982 by Atal and Remde with what was to become known as the Multi-Pulse Excited (MPE) codec. Later the Regular-Pulse Excited (RPE), and the Code-Excited Linear Predictive (CELP) codecs were introduced. These coders will be discussed briefly here.

A general model for AbS codecs is shown in Figure 6.

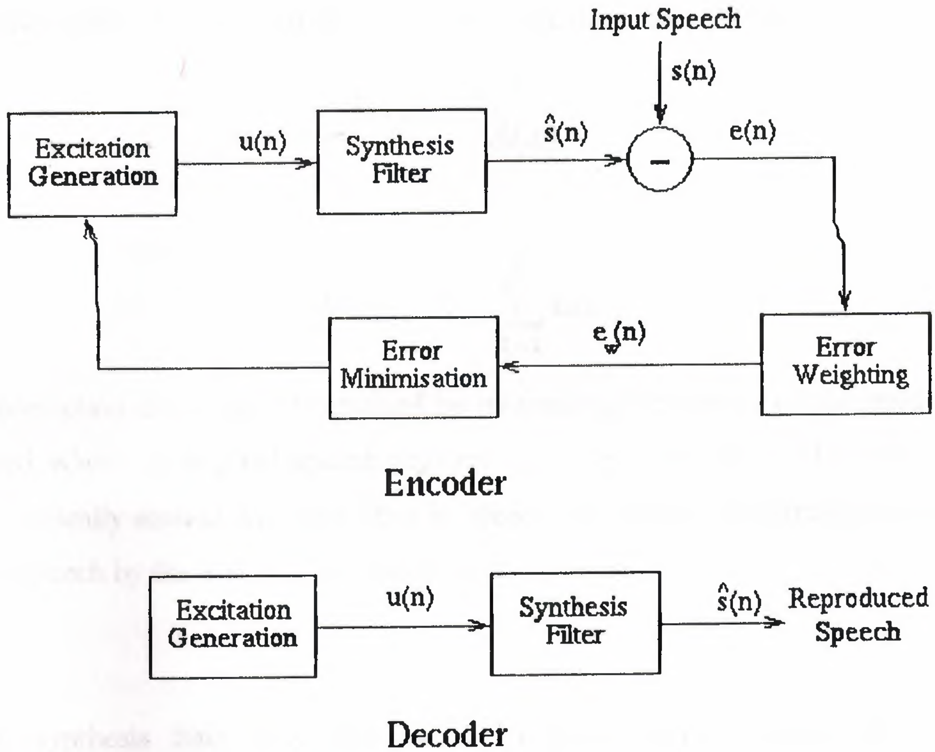

Figure 6: AbS Codec Structure

AbS codecs work by splitting the input speech to be coded into frames, typically about 20 ms long. For each frame parameters are determined for a synthesis filter, and then the excitation to this filter is determined.

This is done by finding the excitation signal which when passed into the given synthesis filter minimes the error between the input speech and the reconstructed speech.

Thus the name Analysis-by-Synthesis - the encoder analyses the input speech by synthesizing many different approximations to it. Finally for each frame the encoder transmits information representing the synthesis filter parameters and the excitation to the decoder, and at the decoder the given excitation is passed through the synthesis filter

The synthesis filter is usually an all pole, short-term, linear filter of the form

$$H(z) = \frac{1}{A(z)}$$

Where :

$$A(z) = 1 - \sum_{i=1}^{p} a_i z^{-i}$$

is the prediction error filter determined by minimizing the energy of the residual signal produced when the original speech segment is passed through it. The order p of the filter is typically around ten. This filter is intended to model the correlations introduced into the speech by the action of the vocal tract.

The synthesis filter may also include a pitch filter to model the long-term periodicities present in voiced speech. Alternatively these long-term periodicities may be exploited by including an adaptive codebook in the excitation generator so that the excitation signal u(n) includes a component of the form Gu(n-α), where α is the estimated pitch period. Generally MPE and RPE codecs will work without a pitch filter, although their performance will be improved if one is included. For CELP codecs however a pitch filter is extremely important, for reasons discussed below.

The error weighting block is used to shape the spectrum of the error signal in order to reduce the subjective loudness of this error. This is possible because the error signal in frequency regions where the speech has high energy will be at least partially masked by the speech. The weighting filter emphasizes the noise in the frequency regions where the speech content is low. Thus minimizing the weighted error concentrates the energy of the error signal in frequency regions where the speech has high energy.

Therefore the error signal will be at least partially masked by the speech, and so its subjective importance will be reduced. Such weighting is found to produce a significant improvement in the subjective quality of the reconstructed speech for AbS codecs.

The distinguishing feature of ABS codecs is how the excitation waveform u(n) for the synthesis filter is chosen. Conceptually every possible waveform is passed through the filter to see what reconstructed speech signal this excitation would produce. The excitation which gives the minimum weighted error between the original and the reconstructed speech is then chosen by the encoder and used to drive the synthesis filter at the decoder. It is this `closed-loop' determination of the excitation which allows ABS codecs to produce good quality speech at low bit rates.

However the numerical complexity involved in passing every possible excitation signal through the synthesis filter is huge. Usually some means of reducing this complexity, without compromising the performance of the codec too badly, must be found.

The differences between MPE, RPE and CELP codecs arise from the representation of the excitation signal u(n) used. In multi-pulse codecs u(n) is given by a fixed number of non-zero pulses for every frame of speech. The positions of these non-zero pulses within the frame, and their amplitudes, must be determined by the encoder and transmitted to the decoder.

In theory it would be possible to find the very best values for all the pulse positions and amplitudes, but this is not practical due to the excessive complexity it would entail.

In practice some sub-optimal method of finding the pulse positions and amplitudes must be used. Typically about 4 pulses per 5 ms are used, and this leads to good quality reconstructed speech at a bit-rate of around 10 kbits/s. Like the MPE codec the Regular Pulse Excited (RPE) codec uses a number of non-zero pulses to give the excitation signal u(n). However in RPE codecs the pulses are regularly spaced at some fixed interval, and the encoder needs only to determine the position of the first pulse and the amplitude of all the pulses.

Therefore less information needs to be transmitted about pulse positions, and so for a given bit rate the RPE codec can use many more non-zero pulses than MPE codecs. For example at a bit rate of about 10 kbits/s around 10 pulses per 5 ms can be used in RPE codecs, compared to 4 pulses for MPE codecs. This allows RPE codecs to give slightly better quality reconstructed speech quality than MPE codecs.

However they also tend to be more complex. The pan-European GSM mobile telephone system uses a simplified RPE codec, with long-term prediction, operating at 13 kbits/s to provide toll quality speech.

Although MPE and RPE codecs can provide good quality speech at rates of around 10 kbits/s and higher, they are not suitable for rates much below this. This is due to the large amount of information that must be transmitted about the excitation pulses' positions and amplitudes. If we attempt to reduce the bit rate by using fewer pulses, or coarsely quantizing their amplitudes, the reconstructed speech quality deteriorates rapidly.

Currently the most commonly used algorithm for producing good quality speech at rates below 10 kbits/s is Code Excited Linear Prediction (CELP). This approach was proposed by Schroeder and Atal in 1985, and differs from MPE and RPE in that the excitation signal is effectively vector quantized. The excitation is given by an entry from a large vector quantizer codebook, and a gain term to control its power.

Typically the codebook index is represented with about 10 bits (to give a codebook size of 1024 entries) and the gain is coded with about 5 bits. Thus the bit rate necessary to transmit the excitation information is greatly reduced - around 15 bits compared to the 47 bits used for example in the GSM RPE codec.

Originally the codebook used in CELP codecs contained white Gaussian sequences. This was because it was assumed that long and short-term predictors would be able to remove nearly all the redundancy from the speech signal to produce a random noise-like residual. Also it was shown that the short-term probability density function (pdf) of this residual was nearly Gaussian.

Schroeder and Atal found that using such a codebook to produce the excitation for long and short-term synthesis filters could produce high quality speech. However to choose which codebook entry to use in an analysis-by-synthesis procedure meant that every excitation sequence had to be passed through the synthesis filters to see how close the reconstructed speech it produced would be to the original. This meant the complexity of the original CELP codec was much too high for it to be implemented in real-time - it took 125 seconds of Cray-1 CPU time to process 1 second of the speech signal. Since 1985 much work on reducing the complexity of CELP codecs, mainly through altering the structure of the codebook, has been done. Also large advances have been made with the speed possible from DSP chips, so that now it is relatively easy to implement a real-time CELP codec on a single, low cost, DSP chip.

The CELP coding principle has been very successful in producing communications to toll quality speech at bit rates between 4.8 and 16 kbits/s. The CCITT standard 16 kbits/s codec produces speech which is almost indistinguishable from 64 kbits/s log-PCM coded speech, while the DoD 4.8 kbits/s codec gives good communications quality speech. Recently much research has been done on codecs operation below 4.8 kbits/s, with the aim being to produce a codec at 2.4 or 3.6 kbits/s with speech quality equivalent to the 4.8 kbits/s DoD CELP. We will briefly describe here a few of the approaches which seem promising in the search for such a codec.

The CELP codec structure can be improved and used at rates below 4.8 kbits/s by classifying speech segments into one of a number of types (for example voiced, unvoiced and transition frames). The different speech segment types are then coded differently with a specially designed encoder for each type. For example for unvoiced frames the encoder will not use any long-term prediction, whereas for voiced frames such prediction is vital but the fixed codebook may be less important.

Such class-dependent codecs have been shown to be capable of producing reasonable quality speech at rates down to 2.4 kbits/s. Multi-Band Excitation (MBE) codecs work by declaring some regions in the frequency domain as voiced and others as unvoiced.

They transmit for each frame a pitch period, spectral magnitude and phase information, and voiced/unvoiced decisions for the harmonics of the fundamental frequency. Originally it was shown that such a structure was capable of producing good quality speech at 8 kbits/s, and since then this rate has been significantly reduced.

Finally Kleijn has suggested an approach for coding voiced segments of speech called Prototype Waveform Interpolation (PWI). This works by sending information about a single pitch cycle every 20-30 ms, and using interpolation to reproduce a smoothly varying quasi-periodic waveform for voiced speech segments. Excellent quality reproduced speech can be obtained for voiced speech at rates as low as 3 kbits/s.

Such a codec can be combined with a CELP type codec for the unvoiced segments to give good quality speech at rates below 4 kbits/s.

2.4 Pulse code modulation (PCM):

Pulse- code modulation (PCM) is the simplest and oldest waveform-coding scheme a pulse-code modulator consists of three basic sections:

1. Sampler.
2. Quantizer.
3. Encoder.

A Functional block diagram of a PCM system is shown in fig.2-2-1.

The waveform entering the sampler is a bandlimited waveform with bandwidths *W*.

usually their exists a filter with bandwidth *W* prior to the sampler to prevent any components beyond *W* from entering the sampler. This filter is called the *presampling filter*. The sampling is done at a rate higher than the Nyquist rate to allow for some guard band. The sampled values then enter scalar quantizer. The quantizer is either uniform quantizer, who results in a uniform PCM system, or a nonuniform quantizer. The choice of the quantizer is based on the characteristics of the source output. The output of the quantizer is then encoded into a binary sequence of length υ where:

$N=2^{\upsilon}$

Are the numbers of quanatization levels.

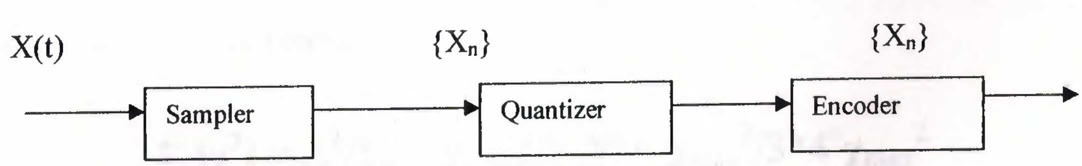

Fig.2-2-1 Block diagram of a PCM system

2.5 UNIFORM (PCM):

In uniform PCM applications, it is assumed that the range of the input samples is $[-\chi_{max}, \chi_{max}]$ and the number of quantization levels N is a power of 2, $N = 2^{\upsilon}$, from this, the length of each quantization region is given by :

$$\Delta = 2\,\chi_{max}/N = \chi_{max}/2^{\upsilon-1}$$

The quantized values in uniform (PCM) are chosen to be the midpoints the quantization regions, and therefore the error:

$$\chi = \chi - Q(\chi)$$

The random variable taken values in the interval $(-\Delta/2, +\Delta/2]$. In ordinary PCM applications, the number of levels (N) is usually high, and the rang of variations of the input signal (amplitude variations χ_{max}) is small.

This means that the length of each quantization region (Δ) is small and, under these assumptions, in each quantization region the error $\chi = \chi - Q(\chi)$ can be well approximated by a uniformly distributed random variable on $(-\Delta/2, +\Delta/2]$.

2.6 Signal to quantization noise ratio (SQNR):

Referred to the previous part. The distortion introduced by quantization (quantization noise) is therefore:

$$E\,[\chi^2] = \Delta^2/12 = \chi_{max}2/3\,N^2 = \chi_{max}{}^2/3*4^{\upsilon}\chi_{max}{}^2$$

Where υ is the number of bits per source sample. The SQNR then becomes:

$$SQNR = \frac{x^2}{x^2} = 3*4^{\upsilon}\chi_{max}^{2}$$

If we denote the normalized X by X, that is , X = X/ χ_{max},

Then:

$$\mathbf{SQNR = 3*4^{\upsilon}X^2}$$

$$\mathbf{|X| \leq 1, X2 \leq 1}$$

Expression the SQNR in dB produces by:

$$\mathbf{SQNR} = p_X + 6\,\upsilon + 4.8$$

It seen that each extra bit (increase in υ by one) increase the SQNR by 6 dB. This increase is comparable to that of an optimal system.

Here we briefly discuss some result concerning the bandwidth requirements of a PCM system. If a signal has bandwidth W, then the minimum number of a sample for perfect reconstruction of the signal is given by sampling theorem and is equal to $2W$ samples/sec. If some guard-band is required, then the number of samples/sec is f_s, which is more than $2W$. For each sample υ bits are used, therefor a total of υf_s bits/sec are Required for transmission of the PCM signal.

In the case of the sampling at the Nyquist rate, this is equal to $2\upsilon W$ bits/sec. The minimum bandwidth requirement for transmission of R bits/sec (or, more precisely, R pulses per second) is R/2 therefore the minimum bandwidth requirement of a PCM is: BW = $\upsilon f_s/2$

Which is the case of sampling at the Nyquist rate gives the absolute minimum bandwidth requirement as:

$$BW = \upsilon W$$

Thus means that a PCM system expands the bandwidth of the original signal by a factor of at least υ.

2.7 Nonuniform PCM:

As long as the statistics of the input signal are close to the uniform distribution, uniform PCM works fine. However, in coding of a certain signal such as speech, the input distribution is a far from being uniformly distributed. For a speech waveform, in particular, there exists a higher probability for smaller amplitudes and lower probability for larger amplitudes.

Therefore, it makes sense to design a quantizer with more quantization regions at lower amplitudes and less quantization regions at larger amplitudes. The resulting quantizer will be a Nonuniform quantizer having quantization regions of various sizes.

first pass the samples through a nonlinear element that compresses the large amplitude (reduce dynamic rang of the signal) and then perform a uniform quantization on the output. The usual method for performing Nonuniform quantization is to At the receiving end, the inverse (expansion) of this nonlinear operation is applied to obtain the sampled value.

This technique is called *companding (compressing expanding)*. A block diagram of this system is shown in fig.2-5-1.

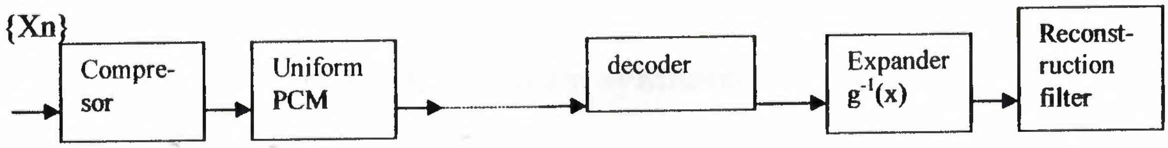

fig.2-5-1 Block diagram of a nonuniform PCM system

There are two types of commanders that are widely used for speech coding. The μ-law commander used in united state and Canada employs the logarithmic function at the transmitting side, with $|x| \leq 1$,

$$g(x) = \log(1+ \mu|x|)*\text{sgn}(x)/\log(1+ \mu).$$

The parameter μ controls the amount of compression. The standard PCM system in the United States and Canada employs a compressor with μ = 255, followed by a uniform quantizer with 128 levels (7 bits/sample). Use of a compander in this system improves the performance of the system by about 24 dB.

The second widely used logarithmic compressor is the A-law compander. It is as follows:

$$g(x) = (1+\ln A|x|)*\text{sgn}(x)/(1+\ln A), \qquad 1/A\leq x\leq 1$$

$$g(x)= A|x|*\text{sgn}(x)/(1+\ln A) \qquad 0 \leq x \leq 1/A$$

3. Spectrum synthesis

3.1 Speech synthesis systems:

Much of interest synthesis system is simulated by the need for economical digital storage of speech for computer voice response system. A computer voice response system is basically in all digital, automatic information service which can be queried by a person from a keyboard or terminal, and which responds with the desired information by voice. Since an ordinary touch-tone telephone can be the keyboard for such a system, the capabilities of such automatic information services can be made universally available over the switched telephone facilities with out the need for any additional specialized equipment.

Speech synthesis systems also play a fundamental role in learning about the process of human speech prediction. Speeches signal processing is very important system in communications. Signal models of speech may operate in either the time or the frequency domain. Since speech is more perceptually relevant in terms of its short-time spectra, rather than its time waveform, a spectral model is particularly convenient the short-time spectra of speech, as exemplified by spectrograms, have defiant structural characteristic that are not to difficult to model.

For voiced speech the spectra are harmonic as voiced-pitch frequency, and are shape by and 'envelope' having three to five resonant peaks (formant) and possibly and anti-resonance.

This type of spectrum can be generated by exciting and appropriate resonant system with a periodic pulse train .for unvoiced speech, the spectra may be characterized as wideband noise, shaped by one or two resonant peaks and one anti-resonance. This may be generated quite simply by exciting and appropriate resonant system with white noise.

The wave shape of the periodic puls is not critical, since its spectrum can be shaped by the resonant system. Likewise, spectral coloring of the white noise source can be taken account of in the design of the resonant system. Control signal for such a basic arrangement are simply: F_o, the fundamental frequency of the pulse source; A_o, its amplitude; An; the amplitude of the noise source; and V_R , a vector which specifies a transfer function of the resonant system.

Depending on the complexity of the synthesiser, V_R may have between 8 & 20 component. The variety of the spectrum modelling techniques described in the literature is due mainly to the variety of ways in which the resonant system may be implemented and controlled. Almost always, periodic pulse and random noise are used as the raw material of the synthetic spectral shapes, as shown in fig.3-1-1.

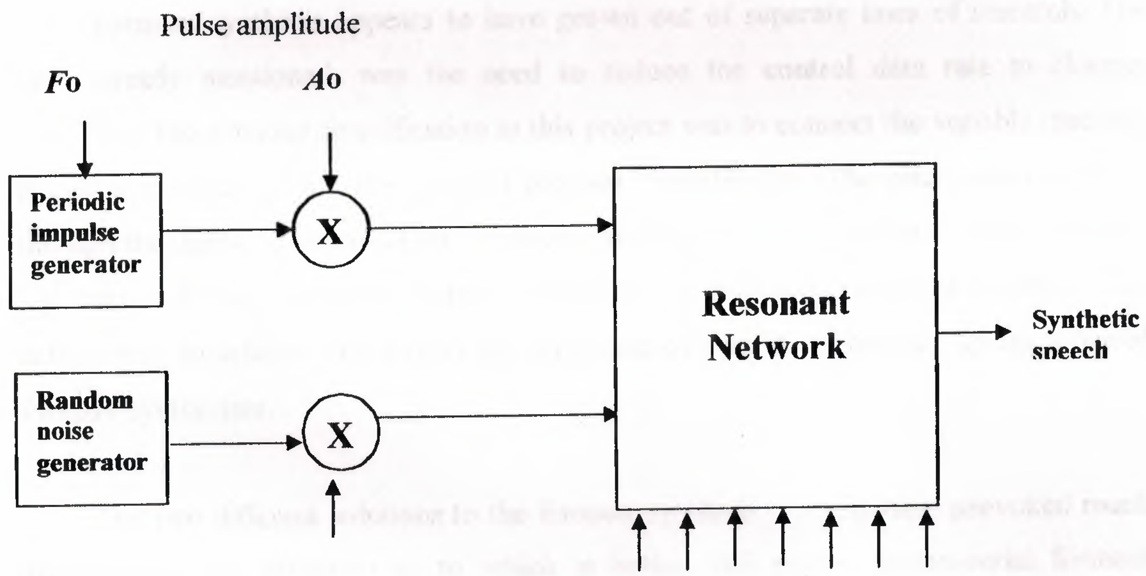

Fig. 3-1-1: Minimal configuration for synthesizing speech like spectra.

3.2 Formant synthesizer:

In the channel synthesizer, controlling the amplitudes of the channel filter outputs creates the three to five formant peaks of the typical speech spectrum.

The twenty or so channels must all be fed with control signals to specify their output amplitude. In the formant synthesizer , only three to five resonant circuits are required, but the resonant frequency of each resonator must be variable as well as the amplitude. This means that fewer control signals are required to specify a particular spectral shape. The exact shape of the spectrum may not be exactly as required, depending on the nature and arrangement of the resonant circuits, but careful design can ensure a close enough 'fit' to give excellent speech quality.

Formant synthesis appears to have grown out of separate lines of research. The first, already mentioned, was the need to reduce the control data rate in channel vocoders. The obvious simplification in this project was to connect the variable resonant circuit in parallel, giving the 'parallel formant ' synthesizer. The other approach was through the desire to construct an 'economy' vocal tract model, without bilateral circuits and with only the minimum number resonators to produce speech like-sounds. The natural way to achieve this way to use a cascade of variable resonators, giving a 'serial formant' synthesiser.

The two different solutions to the formant synthesis problem have provoked much discussion in the literature as to which is better, and parallel-versus-serial formant debate is still a lively topic, even today. A fair exposition of both sides of the argument has worked in both configurations. In summary, the serial type is the better model for non-nasals, fricatives and stops.

The chief merit of the serial formant synthesizer is that its transfer function resembles that of vocal tract (with out nasal coupling), and is therefore good for non-nasal voiced sounds. Usually Only three to five resonators are necessary to obtain acceptable quality speech, thus only three to five formant frequencies and bandwidths need to be specified to define the spectrum. The simplest form of the resonant low pass filter has a function:

$$Z(s) = \frac{\omega_0^2}{s^2 + s\beta + \omega_0^2}$$

where B is the bandwidth, W is the resonant frequency and s is the complex frequency variable. Low pass, second order function of this kind is quite adequate if a compensating network is used to correct for the absence of higher formants. This correction function is illustrated in figure 3-2-1, depend on the number and frequency of the formant used, but in practical situation can be approximated by a fixed network. The fact that formant amplitudes cannot be controlled individually does not matter if the bandwidth is adjusted correctly. The formant amplitude is, in fact, predictable from knowledge of the formant frequencies and bandwidths.

In a practical synthesiser it is not to much of an approximation to fix the bandwidth of each formant so that only three to five formant frequencies are required to specify any non-nasal voiced sound. This makes the serial formant synthesiser very attractive in terms of control data economy.

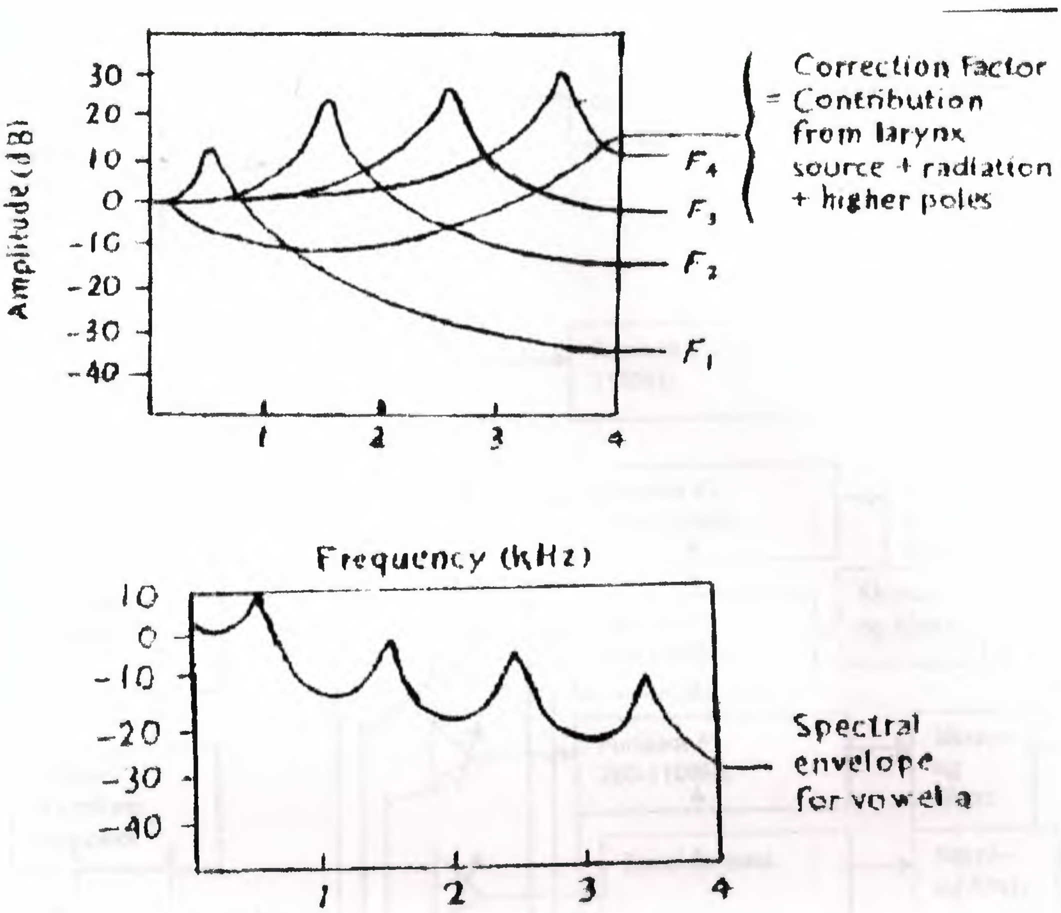

Fig. 3-2-1: Spectrum correction factor for serial formant synthesis.

As an example of a particularly successful design consider figure 3-2-2, which shows the schematic arrangement of the synthesiser. This machine uses four formant resonators connected in parallel. The signs of the formant gains are alternated to minimize the effect of zeros between formants; in addition each formant is followed by spectral shaping filter.

Separate parallel branches are provided for nasals and fricatives. Each formant amplitude is variable, as are the contributions from the nasal and fricative branches. The frequencies of the first three formants can be varied, and also the fundamental frequency of the excitation waveform, which has a variable, pulse widths. The voiced/voiceless option is controlled by a binary switch there are; altogether, twelve control signals when varied properly are capable of giving high quality speech.

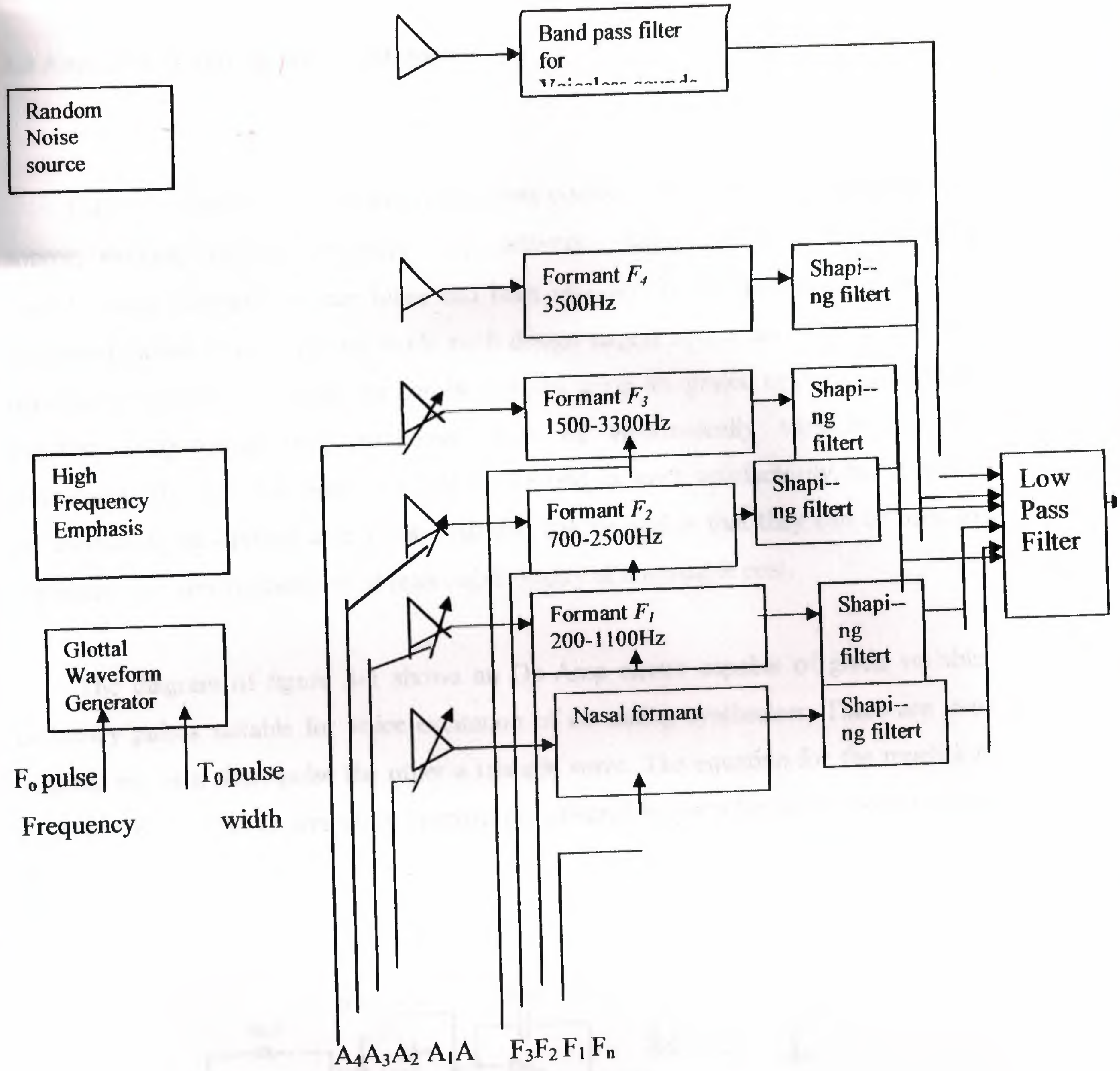

Fig.3-2-2 Schematic diagram of parallel formant synthesizer.

3.3 Analogue circuit formant synthesis:

Circuit to implement formant synthesisers consists essentially of pulse and noise source, exciting variable frequency filter networks. Many ingenious designs using discrete transistors and vacuum tubes had been reported in the literature, but modern integrated circuit technology has made such design largely redundant. All the function required for a formant synthesiser can be realised using integrated circuits operational amplifiers (Op-Amps), together some kind of electronically variable resistive component. The circuit in figure 3-1 had been found to work satisfactorily, but they are not claimed to be optimal in any way. All that can be said is that they can be used to implement formant synthesizers of reasonable quality at reasonable cost.

The diagram of figure 3-1 shows an Op-Amp circuit capable of given variable frequency pulses suitable for voice excitation of an analog synthesiser. There are two outputs, one is a short pulse the other a triangle wave. The equation for the repetition frequency is $F_o = 10^6$ G. frequency controls are affected by variation of the conductance G.

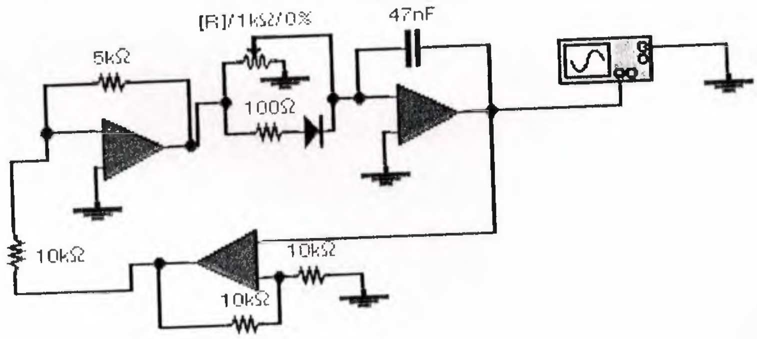

Fig. 3-1: Circuit diagram of variable frequency oscillator for voiced excitation.

After using the oscilloscope to check the circuit, we found that the period of short pulse is equal to 30 μsec as shown in figure 3-1-1 and its data is shown below.

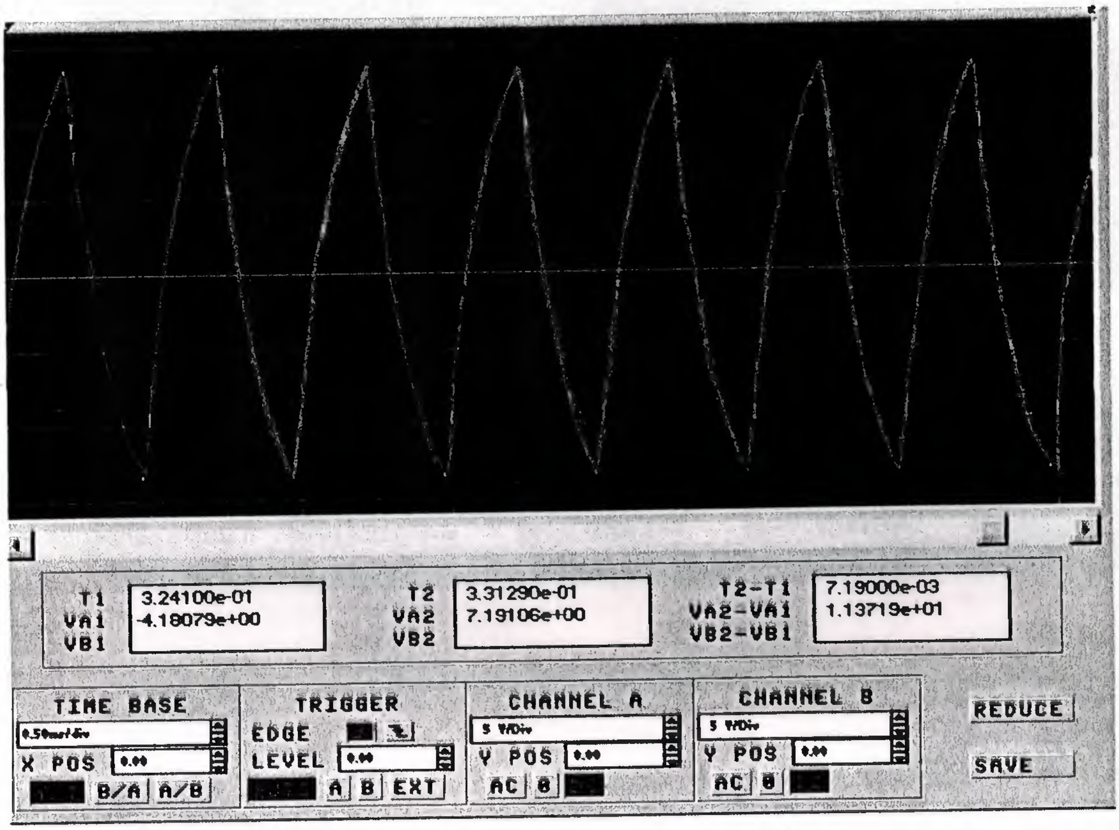

Fig. 3-1-1: waveform short pulses.

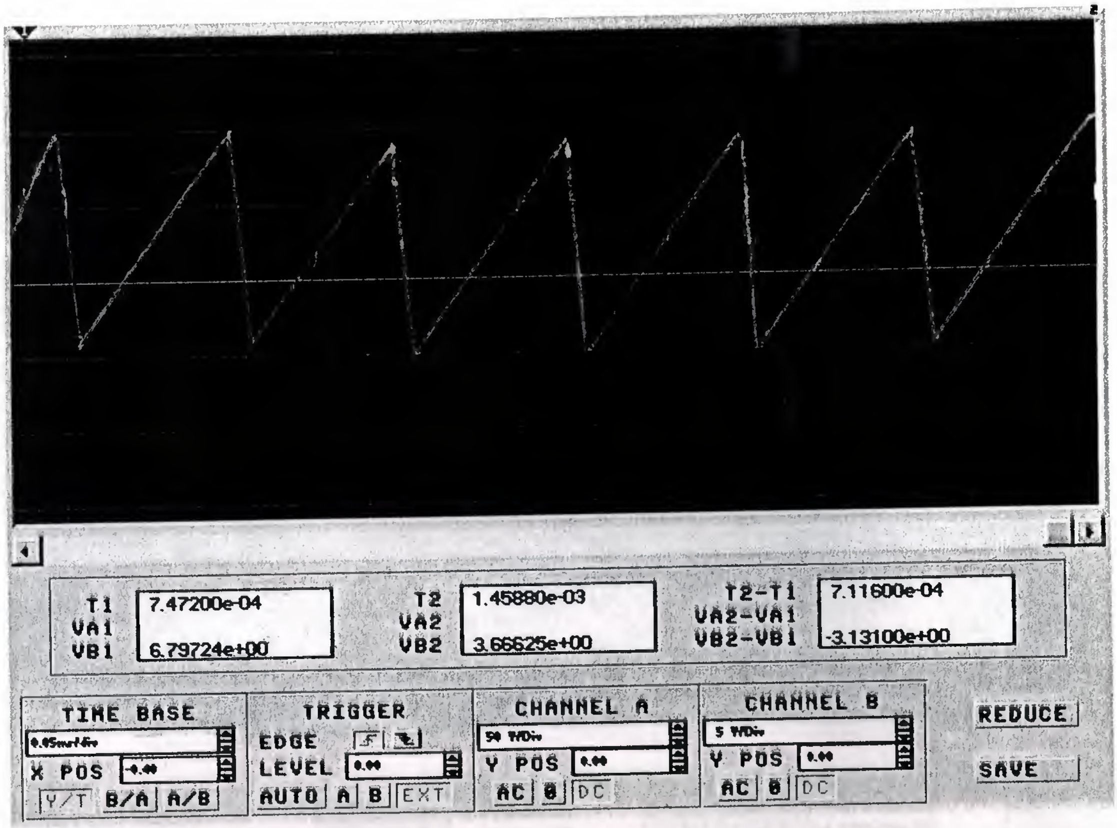

Fig. 3-1-2: Output of the oscillator.

The circuit of figure 3-2 shows random noise generators which uses the reversed bias base emeter junction of a bipolar transistor as the noise source. This gives approximately white noise over a band 100-10,000Hz.

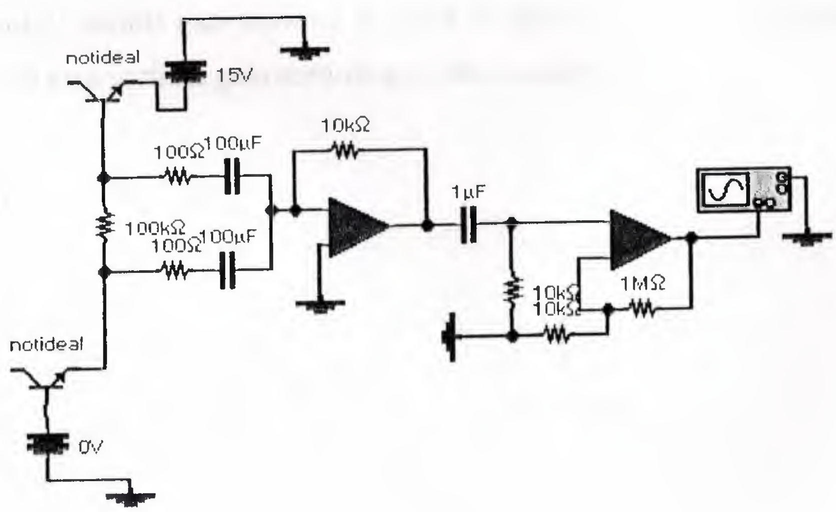

Fig. 3-2: Circuit diagram of white noise generator for unvoiced excitation.

The figure shown below represents the waveform of the white noise generator for unvoiced excitation.

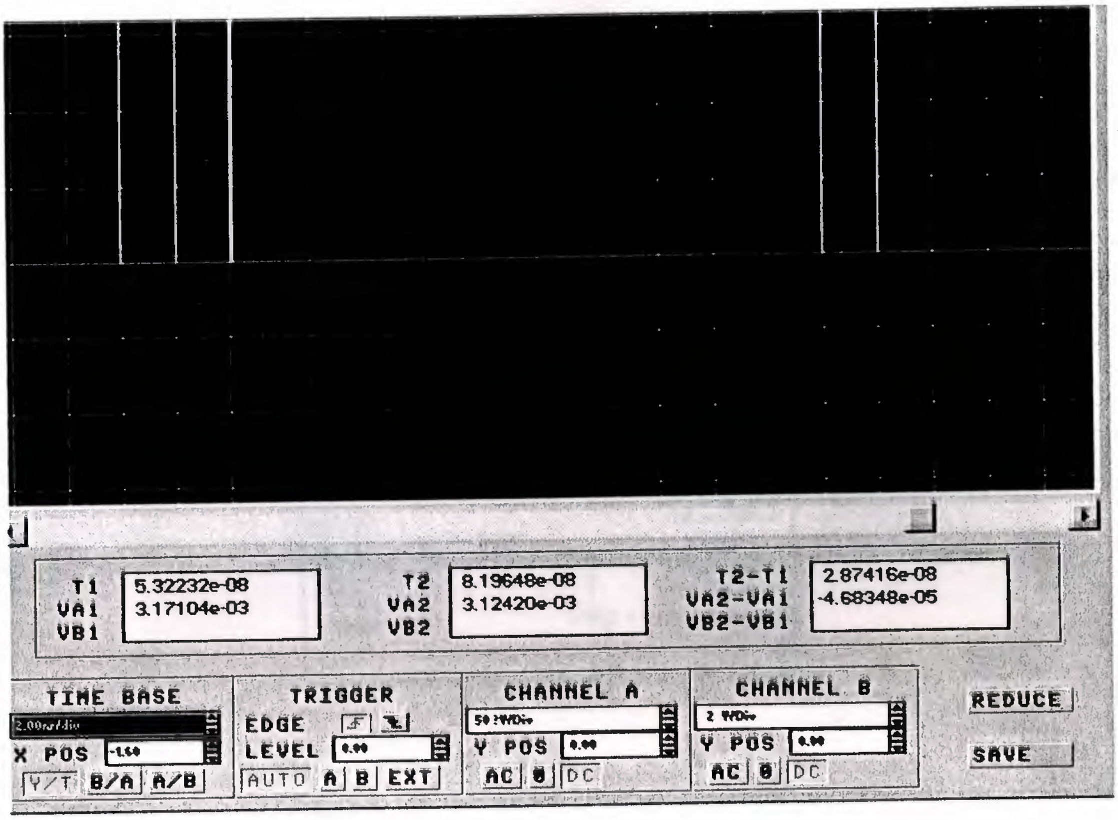

Fig. 3-2-1: The waveform of white noise generator for unvoiced excitation.

A useful variable gain network is given in figure 3-3 variation of either R_1 or R_2 (or both) will give variable gain according to the equation:

$$\mathbf{Gain = - R_2 / R_1}$$

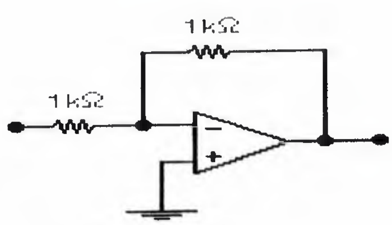

Fig. 3-3: Circuit diagram of variable gain amplifier.

A variable frequency low pass filter network is shown in fig.3-4. This circuit uses variable capacitance to obtain a variation of filter centere frequency. The filter transfer function is given by :

$$Z(s) = (1/LC)/(s^2 + sL/R + 1/LC)$$

The center frequency is given by $(LC)^{-1/2}$ and the bandwidth by R/L.

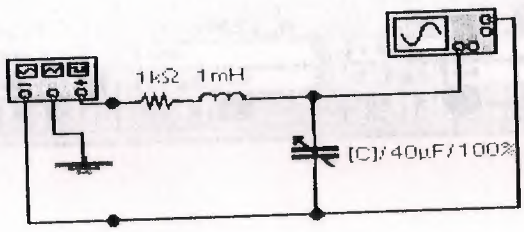

Fig. 3-4: Circuit diagram of serial formant resonator.

Speech signal frequency approximately from 1to 4 KHz , in my design project I choose the input frequency in all circuit equal to 2 KHz.

Calculation:

let C = 1 µF

If input frequency f = 2KHz

R =1 KΩ , L =1mH ,

$f_o = 1/2\pi(LC)^{1/2}$

$f_o = 1/2\pi(10^{-6} * 10^{-3})^{1/2}$

f_o =5.03 KHz

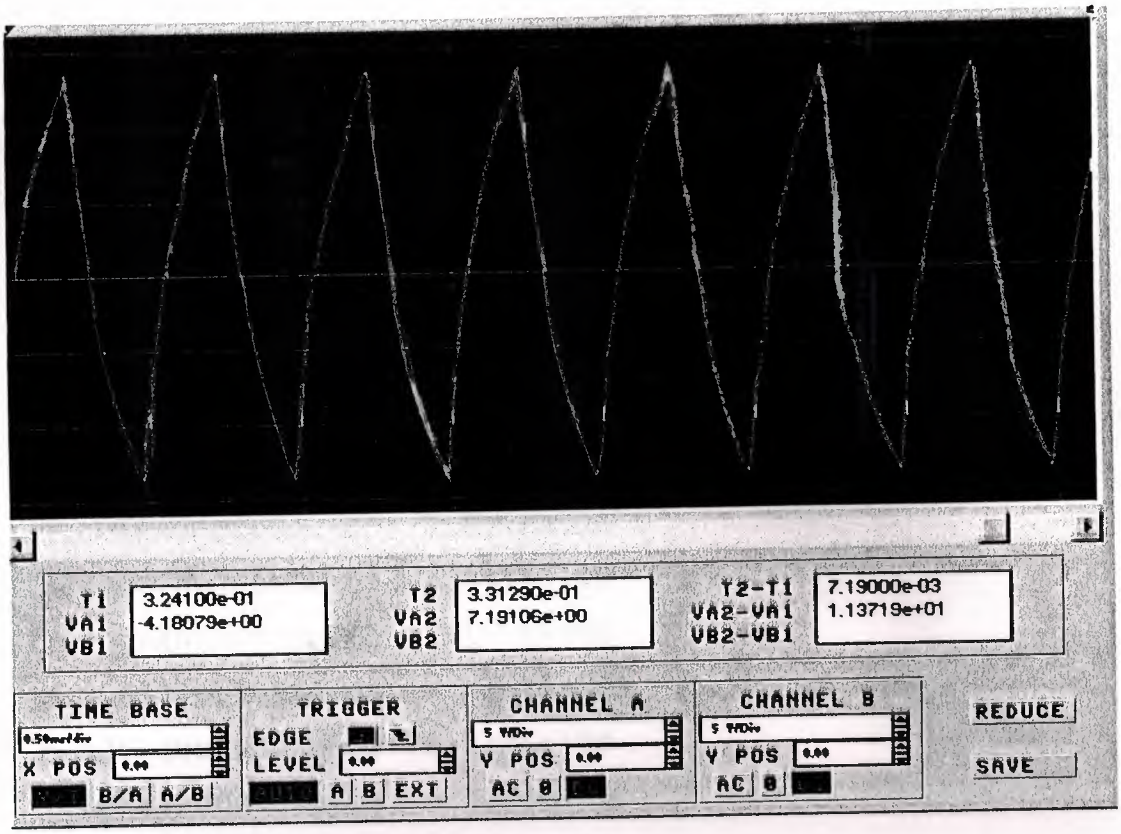

Fig. 3-4-1: The waveform of serial formant resonator with($C = 1\ \mu F$)

Oscilloscope data:

Time base: 0.0005 seconds per division

Time offset: 0 seconds

Channel A sensitivity: 5 volts per division

Channel A offset: 0 volts

Channel B sensitivity: 5 volts per division

Channel B offset: 0 volts

Channel A connected: yes

Channel B connected: no

Column 1 time (seconds)

Column 2 channel A voltage

Time	Channel A
0.000000000000e+00	0.0000e+00
2.000000000000e-09	6.6600e-08
2.200000000000e-09	8.1252e-08
1.000220000000e-05	5.4808e-01
2.000220000000e-05	1.2611e+00
3.000220000000e-05	1.8288e+00
4.000220000000e-05	2.4590e+00
5.000220000000e-05	3.0144e+00
6.000220000000e-05	3.5875e+00
7.000220000000e-05	4.1178e+00
8.000220000000e-05	4.6466e+00
9.000220000000e-05	5.1475e+00
1.000022000000e-04	5.6389e+00
1.100022000000e-04	6.1094e+00
1.200022000000e-04	6.5676e+00
1.300022000000e-04	7.0086e+00
1.400022000000e-04	7.4365e+00
1.500022000000e-04	7.8494e+00
1.600022000000e-04	8.2494e+00
1.700022000000e-04	8.6357e+00
1.800022000000e-04	9.0097e+00
1.900022000000e-04	9.3711e+00
2.000022000000e-04	9.7208e+00
2.100022000000e-04	1.0059e+01
2.200022000000e-04	1.0386e+01
2.300022000000e-04	1.0702e+01
2.400022000000e-04	1.1008e+01
2.500022000000e-04	1.1304e+01
2.600022000000e-04	1.1590e+01
2.700022000000e-04	1.1867e+01
2.800022000000e-04	1.2134e+01
2.900022000000e-04	1.2393e+01

3.000022000000e-04	1.2643e+01
3.100022000000e-04	1.2885e+01
3.200022000000e-04	1.3119e+01
3.300022000000e-04	1.3345e+01
3.400022000000e-04	1.3564e+01
3.500022000000e-04	1.3776e+01
3.600022000000e-04	1.3981e+01
3.700022000000e-04	1.4179e+01
3.800022000000e-04	1.4370e+01
3.900022000000e-04	1.4556e+01
4.000022000000e-04	1.4735e+01
4.100022000000e-04	1.4908e+01

let C = 39.789 μF

$\therefore\ f_o = f = 2$ kHz

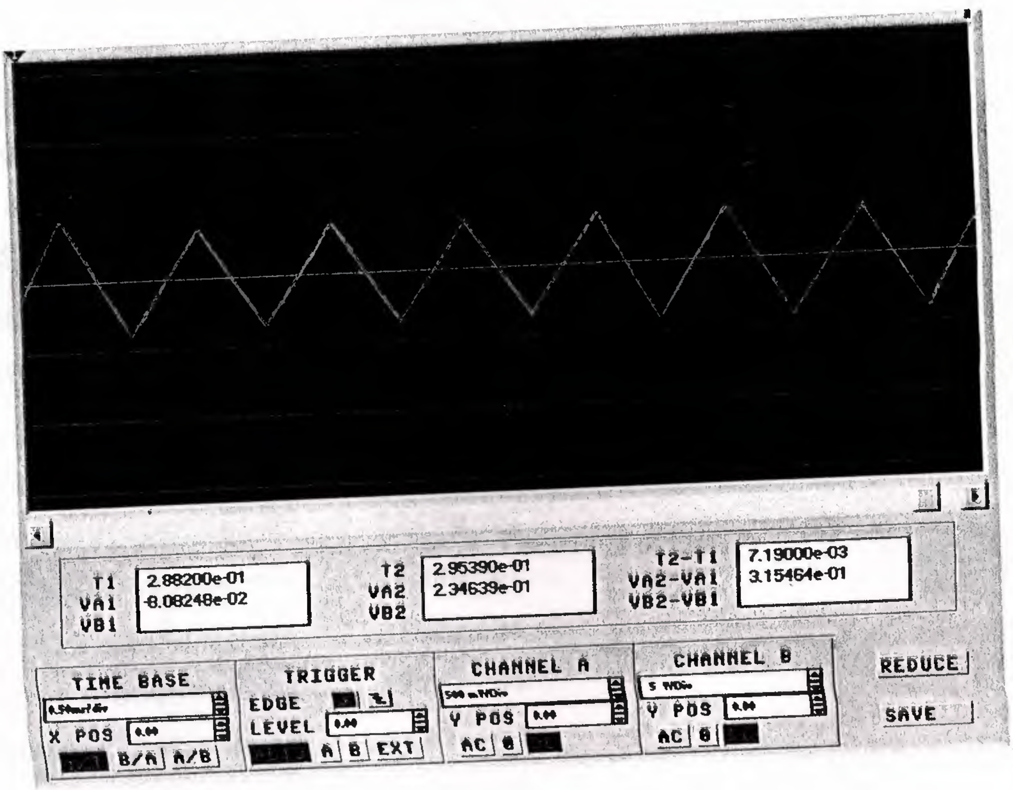

Fig. 3-4-2: The waveform of serial formant resonator with C = 39.789 μF

Oscilloscope data:

Time base: 0.0005 seconds per division
Time offset: 0 seconds
Channel A sensitivity: 0.5 volts per division
Channel A offset: 0 volts
Channel B sensitivity: 5 volts per division
Channel B offset: 0 volts
Channel A connected: yes
Channel B connected: no

Column 1 time (seconds)
Column 2 channel A voltage

Time	Channel A
0.000000000000e+00	0.0000e+00
2.000000000000e-09	1.6650e-09
2.200000000000e-09	2.0313e-09
1.000220000000e-05	1.3887e-02
2.000220000000e-05	3.2386e-02
3.000220000000e-05	4.7787e-02
4.000220000000e-05	6.5231e-02
5.000220000000e-05	8.1289e-02
6.000220000000e-05	9.8248e-02
7.000220000000e-05	1.1458e-01
8.000220000000e-05	1.3131e-01
9.000220000000e-05	1.4776e-01
1.000022000000e-04	1.6437e-01
1.100022000000e-04	1.8084e-01
1.200022000000e-04	1.9738e-01
1.300022000000e-04	2.1386e-01

1.400022000000e-04	2.3036e-01
1.500022000000e-04	2.4682e-01
1.600022000000e-04	2.6328e-01
1.700022000000e-04	2.7972e-01
1.800022000000e-04	2.9615e-01
1.900022000000e-04	3.1256e-01
2.000022000000e-04	3.2896e-01
2.100022000000e-04	3.4535e-01
2.200022000000e-04	3.6172e-01
2.300022000000e-04	3.7808e-01
2.400022000000e-04	3.9443e-01
2.500022000000e-04	4.1076e-01
2.600022000000e-04	4.2708e-01
2.700022000000e-04	4.4338e-01
2.800022000000e-04	4.5968e-01
2.900022000000e-04	4.7595e-01
3.000022000000e-04	4.9222e-01
3.100022000000e-04	5.0847e-01
3.200022000000e-04	5.2471e-01
3.300022000000e-04	5.4093e-01
3.400022000000e-04	5.5714e-01
3.500022000000e-04	5.7334e-01
3.600022000000e-04	5.8952e-01
3.700022000000e-04	6.0569e-01
3.800022000000e-04	6.2185e-01
3.900022000000e-04	6.3799e-01
4.000022000000e-04	6.5412e-01
4.100022000000e-04	6.7024e-01
4.200022000000e-04	6.8634e-01

let C = 10 μF

f_o = 1.59 KHz

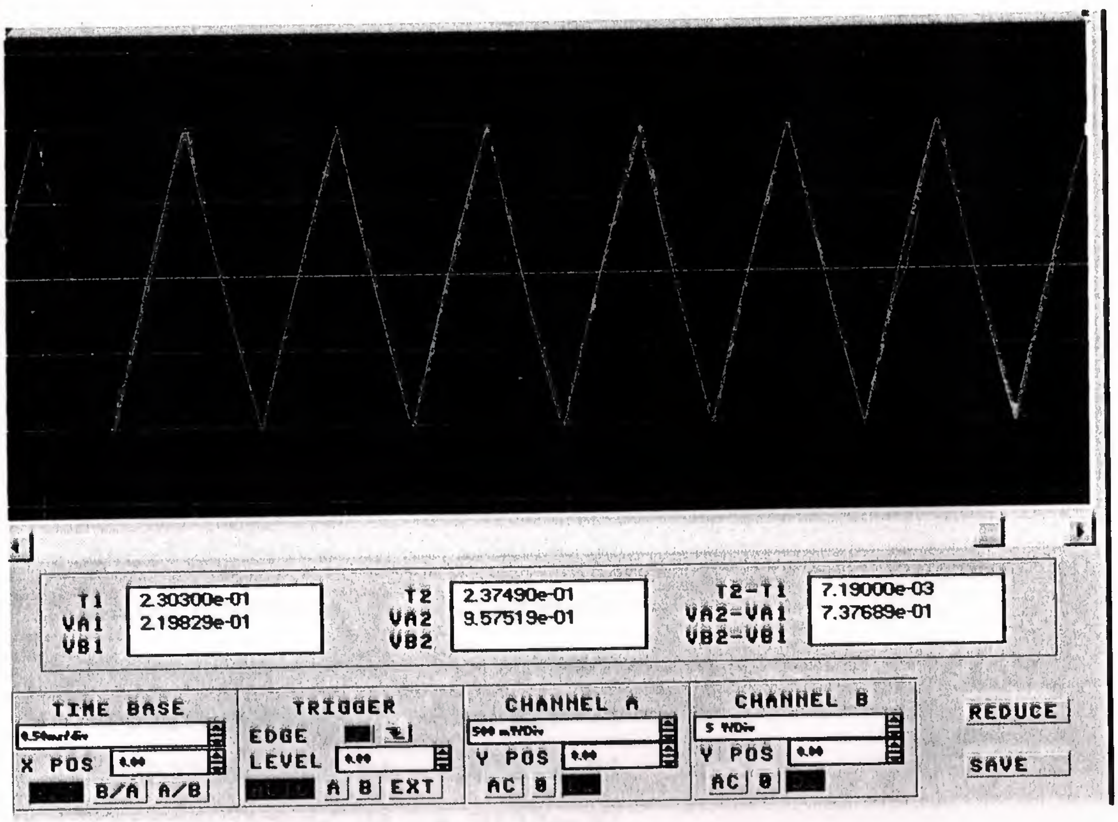

Fig. 3-4-3: The waveform of serial formant resonator C = 10 μF.

Oscilloscope data:

Time base: 0.0005 seconds per division

Time offset: 0 seconds

Channel A sensitivity: 0.5 volts per division

Channel A offset: 0 volts

Channel B sensitivity: 5 volts per division

Channel B offset: 0 volts

Channel A connected: yes

Channel B connected: no

Column 1 time (seconds)

Column 2 channel A voltage

Time	Channel A
0.000000000000e+00	0.0000e+00
2.000000000000e-09	3.9960e-09
2.200000000000e-09	4.8751e-09
1.000220000000e-05	3.3314e-02
2.000220000000e-05	7.7651e-02
3.000220000000e-05	1.1451e-01
4.000220000000e-05	1.5622e-01
5.000220000000e-05	1.9456e-01
6.000220000000e-05	2.3501e-01
7.000220000000e-05	2.7393e-01
8.000220000000e-05	3.1374e-01
9.000220000000e-05	3.5282e-01
1.000022000000e-04	3.9226e-01
1.100022000000e-04	4.3132e-01
1.200022000000e-04	4.7051e-01
1.300022000000e-04	5.0948e-01
1.400022000000e-04	5.4847e-01
1.500022000000e-04	5.8732e-01
1.600022000000e-04	6.2613e-01
1.700022000000e-04	6.6483e-01
1.800022000000e-04	7.0348e-01
1.900022000000e-04	7.4204e-01
2.000022000000e-04	7.8053e-01
2.100022000000e-04	8.1893e-01
2.200022000000e-04	8.5726e-01
2.300022000000e-04	8.9552e-01
2.400022000000e-04	9.3370e-01
2.500022000000e-04	9.7180e-01
2.600022000000e-04	1.0098e+00
2.700022000000e-04	1.0478e+00
2.800022000000e-04	1.0856e+00
2.900022000000e-04	1.1234e+00

3.000022000000e-04	1.1612e+00
3.100022000000e-04	1.1988e+00
3.200022000000e-04	1.2364e+00
3.300022000000e-04	1.2739e+00
3.400022000000e-04	1.3113e+00
3.500022000000e-04	1.3487e+00
3.600022000000e-04	1.3859e+00
3.700022000000e-04	1.4231e+00
3.800022000000e-04	1.4603e+00
3.900022000000e-04	1.4973e+00
4.000022000000e-04	1.5343e+00
4.100022000000e-04	1.5712e+00
4.200022000000e-04	1.6080e+00

The variable capacitance may be realised by using variable feedback as shown in figure 3-5. The effective value of capacitance is C $(R_1 + R_2)/ R_1$. Thus either R_1 or R_2 can be used to control resonant frequency.

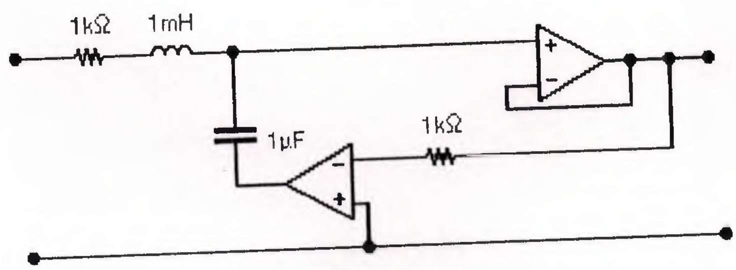

Fig. 3-5: Circuit diagram of variable frequency resonator.

We can replace the variable capacitance in fig. By using a variable feedback.

$$C_{eff} = \frac{R_1 + R_2}{R_1}$$

where C_{eff} : is the variable capacitance

let R = 1 KΩ , L = 1mH

if $R_1 = R_2$ then C_{eff}=2C = 2 μF

∴ $fo = 1 / 2\pi(C_{eff}L) =$ 3.56 KHz

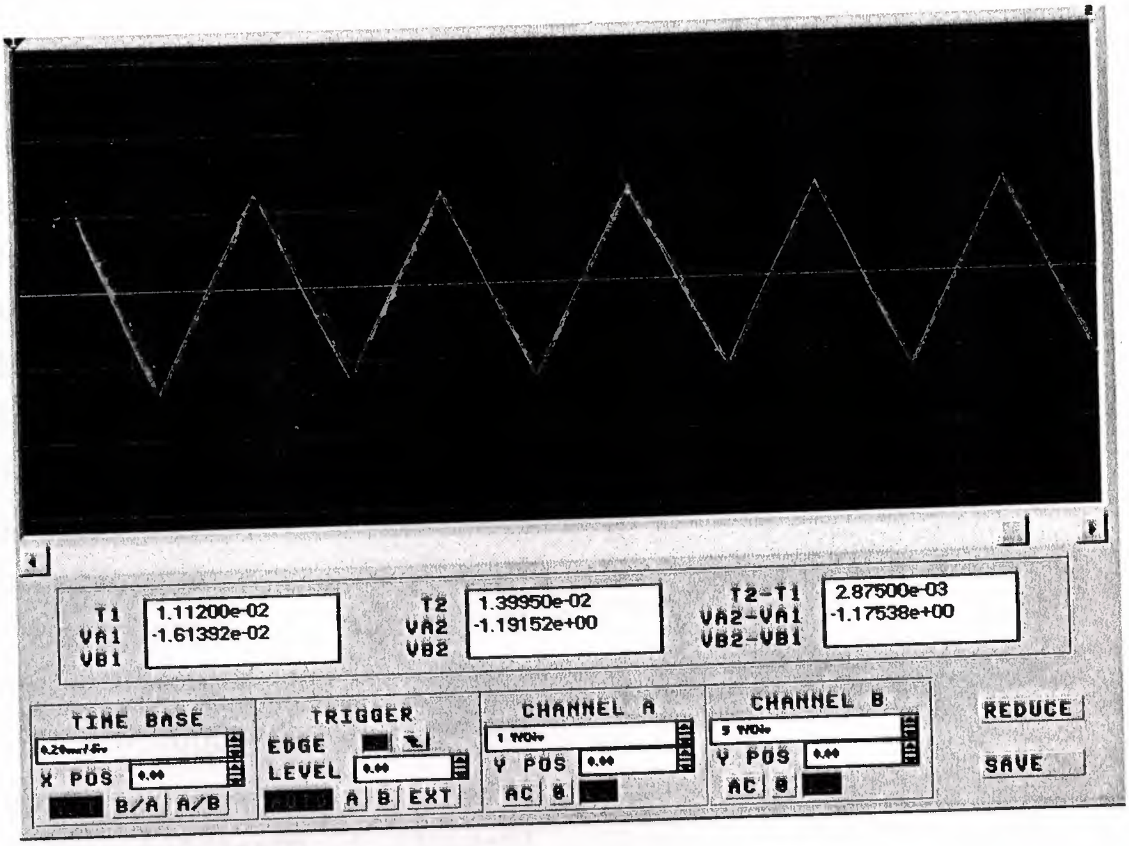

fig. 3-5-1 The waveform of variable frequency resonator with equal R = 1 KΩ.

```
Oscilloscope data:

Time base: 0.0002 seconds per division
Time offset: 0 seconds
Channel A sensitivity: 1 volts per division
Channel A offset: 0 volts
Channel B sensitivity: 5 volts per division
Channel B offset: 0 volts
Channel A connected: yes
Channel B connected: no

Column  1   time (seconds)
Column  2   channel A voltage

Time                    Channel A
```

0.000000000000e+00	0.0000e+00
2.000000000000e-09	1.5010e-05
2.200000000000e-09	1.5012e-05
5.002200000000e-06	3.5715e-02
1.000220000000e-05	9.1698e-02
1.500220000000e-05	1.3881e-01
2.000220000000e-05	1.8955e-01
2.500220000000e-05	2.3856e-01
3.000220000000e-05	2.8813e-01
3.500220000000e-05	3.3729e-01
4.000220000000e-05	3.8644e-01
4.500220000000e-05	4.3543e-01
5.000220000000e-05	4.8431e-01
5.500220000000e-05	5.3306e-01
6.000220000000e-05	5.8169e-01
6.500220000000e-05	6.3020e-01
7.000220000000e-05	6.7859e-01
7.500220000000e-05	7.2685e-01
8.000220000000e-05	7.7500e-01
8.500220000000e-05	8.2303e-01
9.000220000000e-05	8.7093e-01
9.500220000000e-05	9.1872e-01
1.000022000000e-04	9.6639e-01
1.050022000000e-04	1.0139e+00
1.100022000000e-04	1.0614e+00
1.150022000000e-04	1.1087e+00
1.200022000000e-04	1.1559e+00
1.250022000000e-04	1.2029e+00

if $R_1 = 10R_2$ then $C_{eff} = 1.1\ C = 1.1\ \mu F$

$\therefore f_o = 1 / 2\pi(C_{eff} L) = 4.80$ KHz

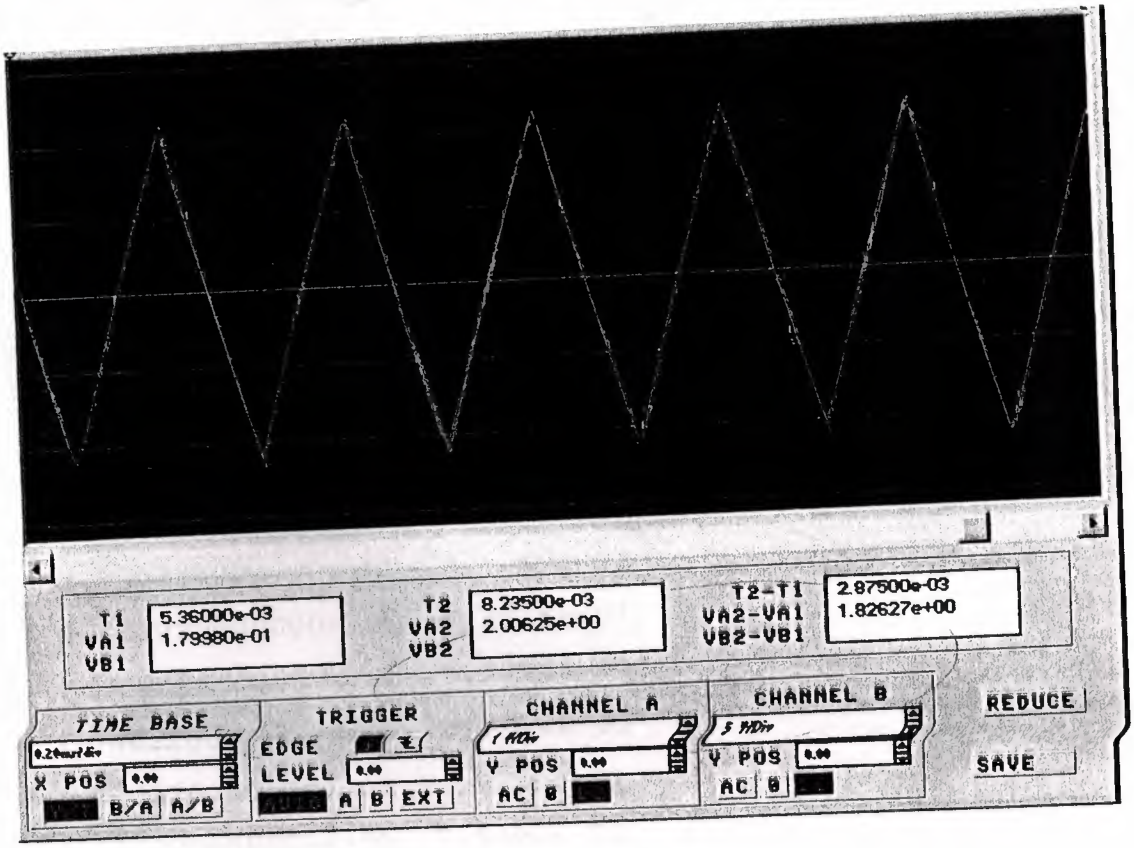

figure 3-5-2 The waveform of variable frequency resonator with R1 = 10R2 .

```
Oscilloscope data:

Time base: 0.0002 seconds per division
Time offset: 0 seconds
Channel A sensitivity: 1 volts per division
Channel A offset: 0 volts
Channel B sensitivity: 5 volts per division
Channel B offset: 0 volts
Channel A connected: yes
Channel B connected: no
```

```
Column  1   time (seconds)
Column  2   channel A voltage
```

Time	Channel A
0.000000000000e+00	0.0000e+00
2.000000000000e-09	1.9109e-05
2.200000000000e-09	1.9113e-05
5.002200000000e-06	6.4880e-02
1.000220000000e-05	1.6647e-01
1.500220000000e-05	2.5175e-01
2.000220000000e-05	3.4343e-01
2.500220000000e-05	4.3178e-01
3.000220000000e-05	5.2098e-01
3.500220000000e-05	6.0924e-01
4.000220000000e-05	6.9733e-01
4.500220000000e-05	7.8492e-01
5.000220000000e-05	8.7215e-01
5.500220000000e-05	9.5897e-01
6.000220000000e-05	1.0454e+00
6.500220000000e-05	1.1314e+00
7.000220000000e-05	1.2171e+00
7.500220000000e-05	1.3024e+00
8.000220000000e-05	1.3872e+00
8.500220000000e-05	1.4717e+00
9.000220000000e-05	1.5558e+00
9.500220000000e-05	1.6395e+00
1.000022000000e-04	1.7229e+00
1.050022000000e-04	1.8059e+00
1.100022000000e-04	1.8884e+00
1.150022000000e-04	1.9707e+00
1.200022000000e-04	2.0525e+00

1.250022000000e-04	2.1340e+00
1.300022000000e-04	2.2151e+00
1.350022000000e-04	2.2958e+00
1.400022000000e-04	2.3762e+00

if $R_2 = 10R_1$ then $C_{eff} = 11C = 11\ \mu F$

$\therefore f_o = 1 / 2\pi(C_{eff} L) = 1.52$ KHz

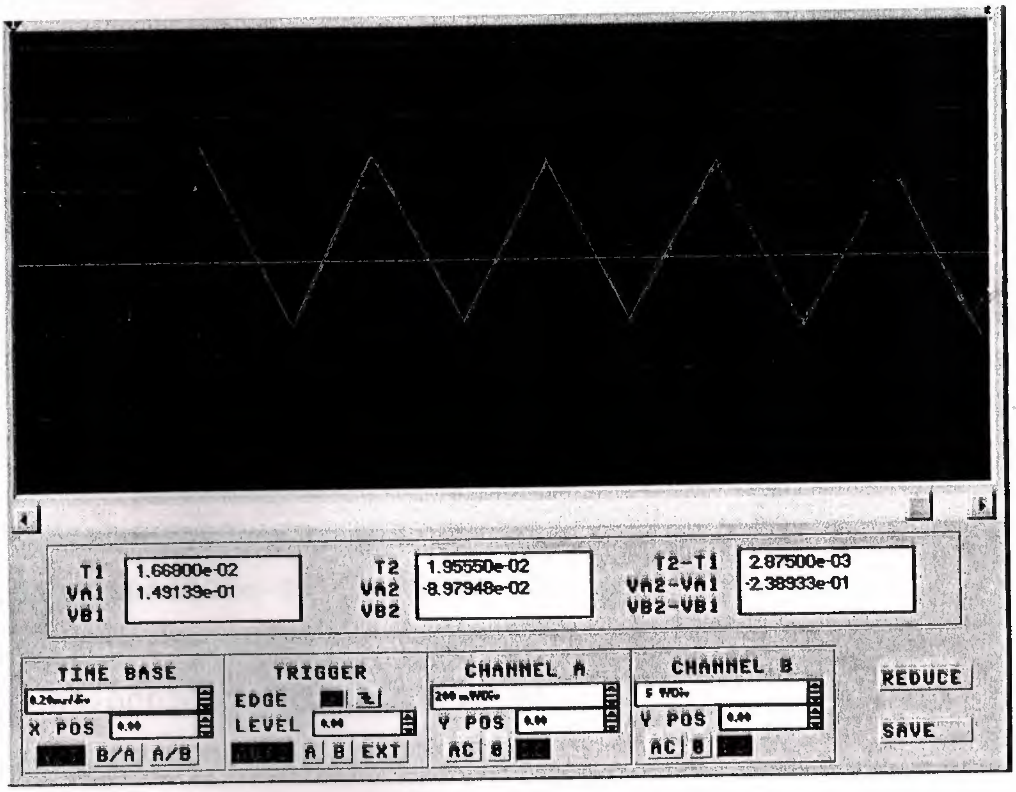

figure 3-5-3 The waveform of variable frequency resonator with R2= 10R1.

Oscilloscope data:

Time base: 0.0002 seconds per division

Time offset: 0 seconds

Channel A sensitivity: 0.2 volts per division

Channel A offset: 0 volts

Channel B sensitivity: 5 volts per division

Channel B offset: 0 volts

Channel A connected: yes

Channel B connected: no

Column 1 time (seconds)

Column 2 channel A voltage

Time	Channel A
0.000000000000e+00	0.0000e+00
2.000000000000e-09	1.0911e-05
2.200000000000e-09	1.0911e-05
5.002200000000e-06	6.5066e-03
1.000220000000e-05	1.6705e-02
1.500220000000e-05	2.5311e-02
2.000220000000e-05	3.4593e-02
2.500220000000e-05	4.3580e-02
3.000220000000e-05	5.2687e-02
3.500220000000e-05	6.1737e-02
4.000220000000e-05	7.0806e-02
4.500220000000e-05	7.9860e-02
5.000220000000e-05	8.8915e-02
5.500220000000e-05	9.7964e-02
6.000220000000e-05	1.0701e-01
6.500220000000e-05	1.1605e-01
7.000220000000e-05	1.2509e-01
7.500220000000e-05	1.3412e-01
8.000220000000e-05	1.4315e-01
8.500220000000e-05	1.5217e-01
9.000220000000e-05	1.6119e-01
9.500220000000e-05	1.7021e-01
1.000022000000e-04	1.7922e-01
1.050022000000e-04	1.8823e-01
1.100022000000e-04	1.9724e-01

1.150022000000e-04	2.0624e-01
1.200022000000e-04	2.1523e-01
1.250022000000e-04	2.2422e-01
1.300022000000e-04	2.3321e-01
1.350022000000e-04	2.4220e-01
1.400022000000e-04	2.5118e-01
1.450022000000e-04	2.6015e-01
1.500022000000e-04	2.6912e-01
1.550022000000e-04	2.7809e-01
1.600022000000e-04	2.8705e-01

Band pass resonators and anti-resonators are best realised using active filter circuit. Figure 3-6 shows general-purpose circuit using four Op-Amps which are capable of given band pass low pass and anti resonant responses.

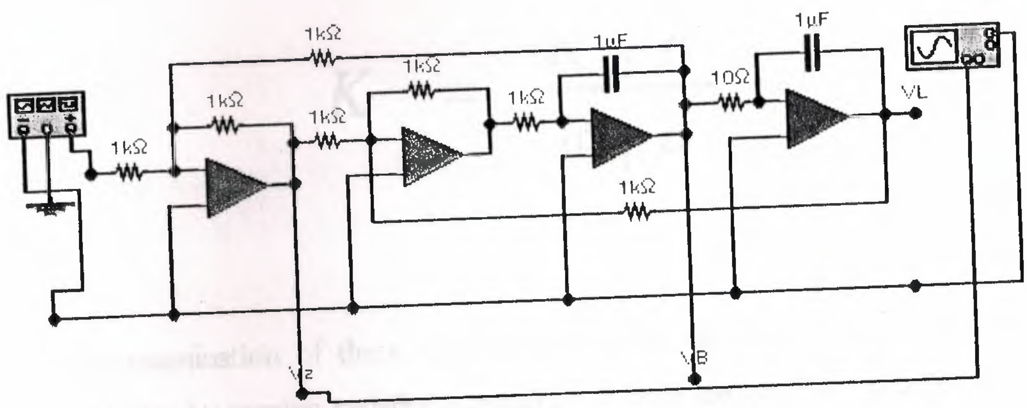

Fig. 3-6: Circuit diagram of general-purpose active filter, suitable for formant synthesis. VL is low pass output. VB is band pass output. VZ is transmission zero output.

Both center frequency and bandwidth are variable by variation of resistive components. A low pass resonance, specified by:

Is available from the output of the fourth amplifier

Where ω_o is the resonant frequency, β is the bandwidth and k is the gain factor. These parameters, given in terms of the circuit component are as follows:

$$\omega_o^{\ 2} = \frac{R_4}{R_5 R_7 R_8 C^2}$$

$$\beta = \frac{R_2 R_4}{R_3 R_6 R_7 C}$$

$$K = \frac{R_2 R_5}{R_1 R_3}$$

An examination of these expressions reveals that ωo, β and k may be varied independently by varying resistive components. Thus ωo may be varied by varying R8, k by varying R1, and β by R6.

A band pass output is available from the third amplifier, giving the response:

$$\frac{V_B}{V_1} = \frac{K \beta s}{s^2 + s\beta + \omega_o^{\ 2}}$$

Where ω_o is the resonant frequency, β is the bandwidth and k is the gain factor. These parameters, given in terms of the circuit component are as follows:

$$\omega_o^{\ 2} = \frac{R_4}{R_5 R_7 R_8 C^2}$$

$$\beta = \frac{R_2 R_4}{R_3 R_6 R_7 C}$$

$$K = \frac{R_2 R_5}{R_1 R_3}$$

An examination of these expressions reveals that ωo, β and k may be varied independently by varying resistive components. Thus ωo may be varied by varying R8, k by varying R1, and β by R6.

A band pass output is available from the third amplifier, giving the response:

$$\frac{V_B}{V_1} = \frac{K\beta s}{s^2 + s\beta + \omega_o^{\ 2}}$$

In this case k=-R6/R1, whilst center frequency and bandwidth are the same as before. A gain, ωo, β and k can be varied independently. ωo Is varied by variaton of R5 and/orR8 , β by R3 ,and k by R1 .

An anti-resonance response is available from the output of the first amplifier, of the type:

$$\frac{V_z}{V_1} = \frac{k\,(s^2 + \omega_o^{\,2})}{s^2 + s\beta + \omega_o^{\,2}}$$

This gives a real frequency transmission zero at ωo. In this case:

K=-R6/R1

Center frequency ωo and bandwidth β are the same as the above. Ωo may be varied using R5 and/or R8, β by R3 and/or R6,and k by R1.

Input frequency f = 2 KHz and all resistance (R1.......R8) are equals to 1 KΩ

And from equation as follows:

$$\omega_o^{\,2} = \frac{R_4}{R_5 R_7 R_8 C^2}$$

$\therefore f_o = (R_4)^2 / (R_5R_7R_8C^2)^2$

If C = 0.1 μF

$\therefore f_o$ = 1.6 KHz, β = 10 KHz

For low pass resonance k = 1

And for band pass k = -1

And for anti-resonance k = -1

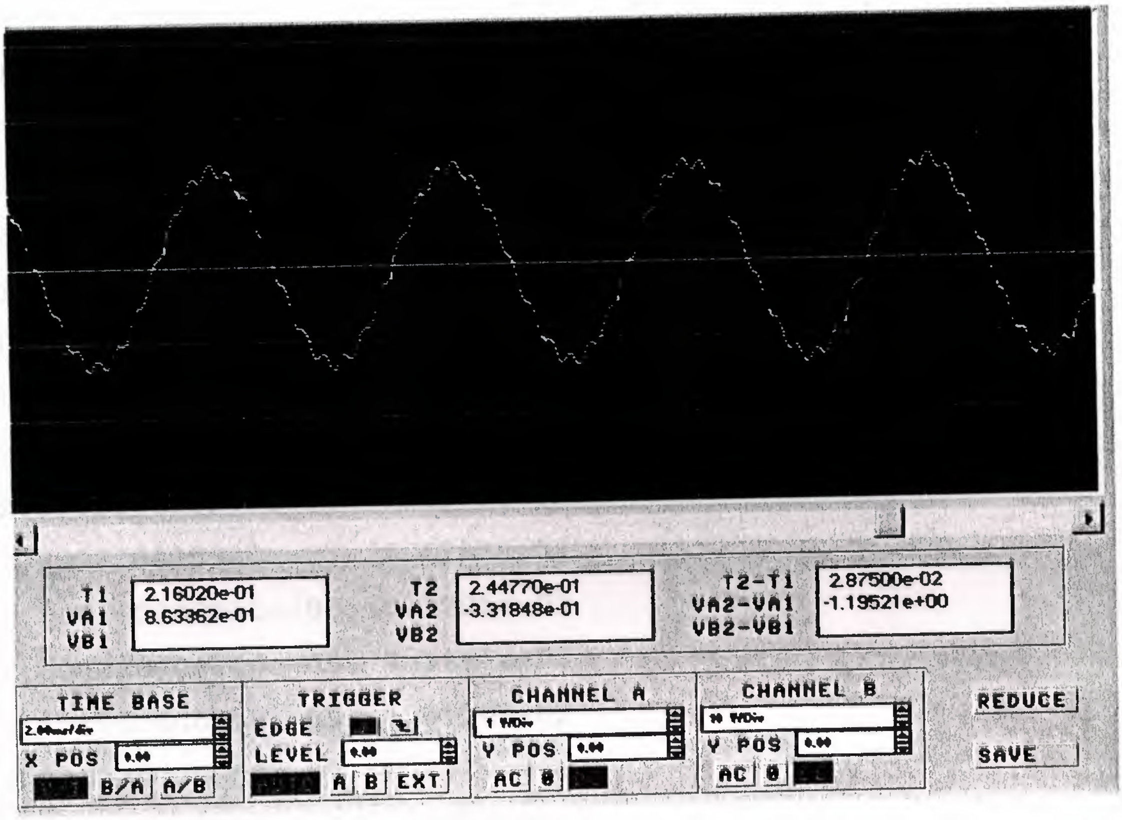

Fig. 3-6-1: The waveform of the low pass output (VL) with $R_8=1\Omega$.

Oscilloscope data:

Time base: 0.002 seconds per division

Time offset: 0 seconds

Channel sensitivity: 1 volt per division

Channel An offset: 0 volts

Channel B sensitivity: 10 volts per division

Channel B offset: 0 volts

Channel A connected: yes

Channel B connected: no

Column 1 time (seconds)

Column 2 channel A voltage

Time	Channel A
0.000000000000e+00	0.0000e+00
2.000000000000e-09	1.0000e-05
2.100000000000e-09	1.0000e-05
2.200000000000e-09	1.0000e-05
5.002200000000e-06	1.3477e-04
1.000220000000e-05	5.0895e-04
1.500220000000e-05	1.1325e-03
2.000220000000e-05	2.0055e-03
2.500220000000e-05	3.1278e-03
3.000220000000e-05	4.4995e-03
3.500220000000e-05	6.1205e-03
4.000220000000e-05	7.9907e-03
4.500220000000e-05	1.0110e-02
5.000220000000e-05	1.2479e-02
5.500220000000e-05	1.5097e-02
6.000220000000e-05	1.7963e-02
6.500220000000e-05	2.1079e-02
7.000220000000e-05	2.4444e-02
7.500220000000e-05	2.8057e-02
8.000220000000e-05	3.1919e-02
8.500220000000e-05	3.6030e-02
9.000220000000e-05	4.0390e-02
9.500220000000e-05	4.4997e-02
1.000022000000e-04	4.9853e-02
1.050022000000e-04	5.4958e-02
1.100022000000e-04	6.0310e-02
1.150022000000e-04	6.5910e-02
1.200022000000e-04	7.1758e-02

1.250022000000e-04	7.7854e-02
1.300022000000e-04	8.4197e-02
1.350022000000e-04	9.0787e-02
1.400022000000e-04	9.7625e-02
1.450022000000e-04	1.0471e-01
1.500022000000e-04	1.1204e-01
1.550022000000e-04	1.1962e-01
1.600022000000e-04	1.2744e-01
1.650022000000e-04	1.3551e-01
1.700022000000e-04	1.4383e-01
1.750022000000e-04	1.5239e-01
1.800022000000e-04	1.6120e-01
1.850022000000e-04	1.7025e-01
1.900022000000e-04	1.7955e-01
1.950022000000e-04	1.8910e-01

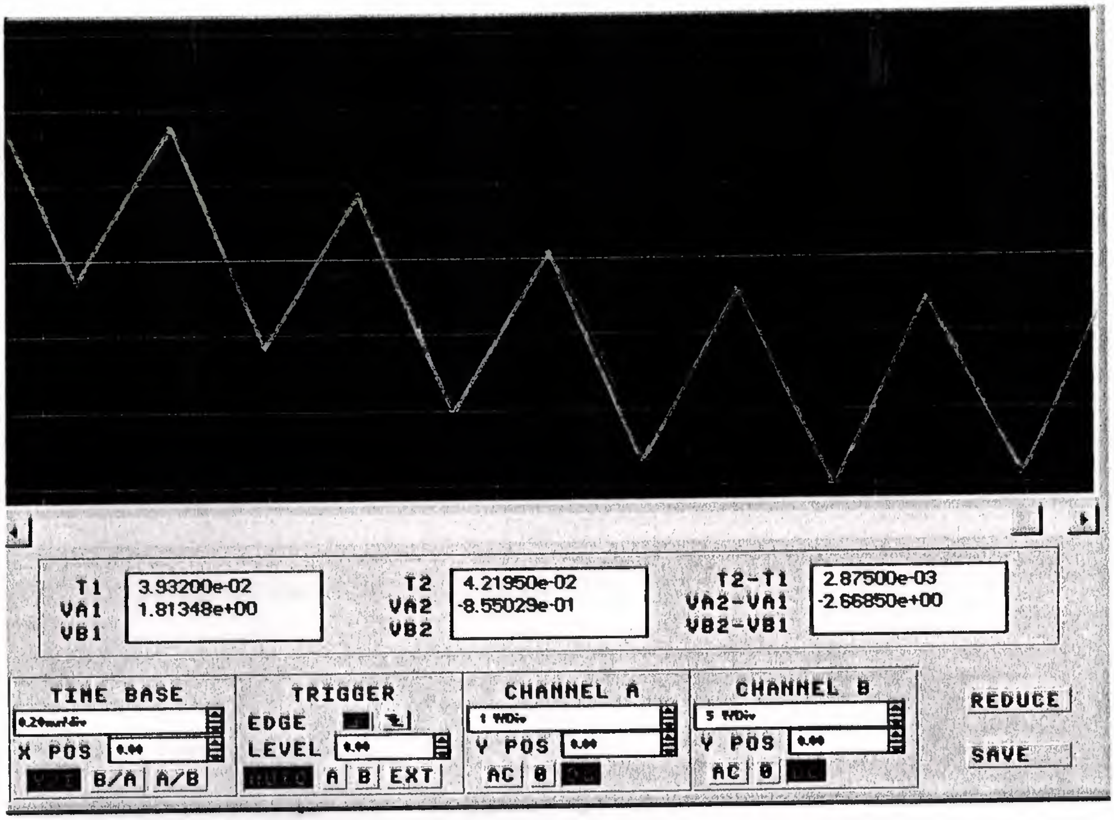

Fig. 3-6-2: The waveform of the band pass output (V_B).

```
Oscilloscope data:

Time base: 0.0002 seconds per division
Time offset: 0 seconds
Channel sensitivity: 1 volt per division
Channel A offset: 0 volts
Channel B sensitivity: 5 volts per division
Channel B offset: 0 volts
Channel A connected: yes
Channel B connected: no

Column  1   time (seconds)
Column  2   channel A voltage

Time                    Channel A
-------------------------------
 0.000000000000e+00  0.0000e+00
```

2.000000000000e-09	9.9634e-06
2.100000000000e-09	1.0961e-05
2.200000000000e-09	1.1959e-05
5.002200000000e-06	5.0020e-02
1.000220000000e-05	1.0028e-01
1.500220000000e-05	1.5078e-01
2.000220000000e-05	2.0153e-01
2.500220000000e-05	2.5252e-01
3.000220000000e-05	3.0376e-01
3.500220000000e-05	3.5524e-01
4.000220000000e-05	4.0696e-01
4.500220000000e-05	4.5891e-01
5.000220000000e-05	5.1111e-01
5.500220000000e-05	5.6354e-01
6.000220000000e-05	6.1621e-01
6.500220000000e-05	6.6911e-01
7.000220000000e-05	7.2225e-01
7.500220000000e-05	7.7562e-01
8.000220000000e-05	8.2921e-01
8.500220000000e-05	8.8304e-01
9.000220000000e-05	9.3709e-01
9.500220000000e-05	9.9137e-01
1.000022000000e-04	1.0459e+00

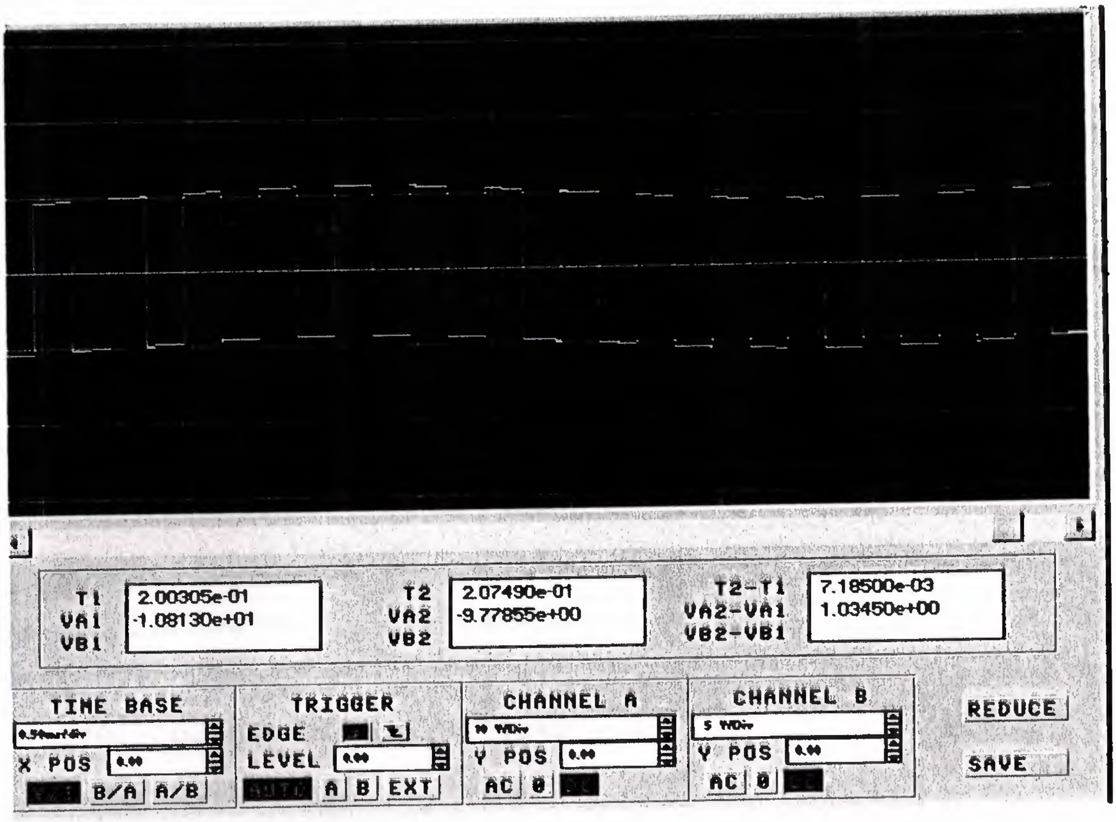

Fig. 3-6-3: The waveform of the anti-resonant output (Vz).

```
Oscilloscope data:

Time base: 0.0005 seconds per division
Time offset: 0 seconds
Channel A sensitivity: 10 volts per division
Channel A offset: 0 volts
Channel B sensitivity: 5 volts per division
Channel B offset: 0 volts
Channel A connected: yes
Channel B connected: no

Column  1   time (seconds)
Column  2   channel A voltage

Time                    Channel A
---------------------------------
 0.000000000000e+00   0.0000e+00
```

2.000000000000e-09	9.9767e+00
2.100000000000e-09	9.9767e+00
2.200000000000e-09	9.9767e+00
5.002200000000e-06	9.9769e+00
1.000220000000e-05	9.9772e+00
1.500220000000e-05	9.9779e+00
2.000220000000e-05	9.9788e+00
2.500220000000e-05	9.9799e+00
3.000220000000e-05	9.9813e+00
3.500220000000e-05	9.9829e+00
4.000220000000e-05	9.9848e+00
4.500220000000e-05	9.9870e+00
5.000220000000e-05	9.9894e+00
5.500220000000e-05	9.9920e+00
6.000220000000e-05	9.9949e+00
6.500220000000e-05	9.9980e+00
7.000220000000e-05	1.0001e+01
7.500220000000e-05	1.0005e+01
8.000220000000e-05	1.0009e+01
8.500220000000e-05	1.0013e+01
9.000220000000e-05	1.0017e+01
9.500220000000e-05	1.0022e+01
1.000022000000e-04	1.0027e+01
1.050022000000e-04	1.0032e+01
1.100022000000e-04	1.0037e+01
1.150022000000e-04	1.0043e+01
1.200022000000e-04	1.0049e+01
1.250022000000e-04	1.0055e+01
1.300022000000e-04	1.0061e+01
1.350022000000e-04	1.0068e+01
1.400022000000e-04	1.0075e+01
1.450022000000e-04	1.0082e+01
1.500022000000e-04	1.0089e+01
1.550022000000e-04	1.0097e+01

1.600022000000e-04	1.0105e+01
1.650022000000e-04	1.0113e+01
1.700022000000e-04	1.0121e+01
1.750022000000e-04	1.0130e+01
1.800022000000e-04	1.0139e+01
1.850022000000e-04	1.0148e+01
1.900022000000e-04	1.0157e+01
1.950022000000e-04	1.0166e+01

Since R8 is command between three cases (low pass, band pass and anti-resonance) I change R8 value to 10 kΩ , and (R1.......R7) are constant, that the center frequency is change as follow: .

$\therefore f_o = 0.503$ KΩ

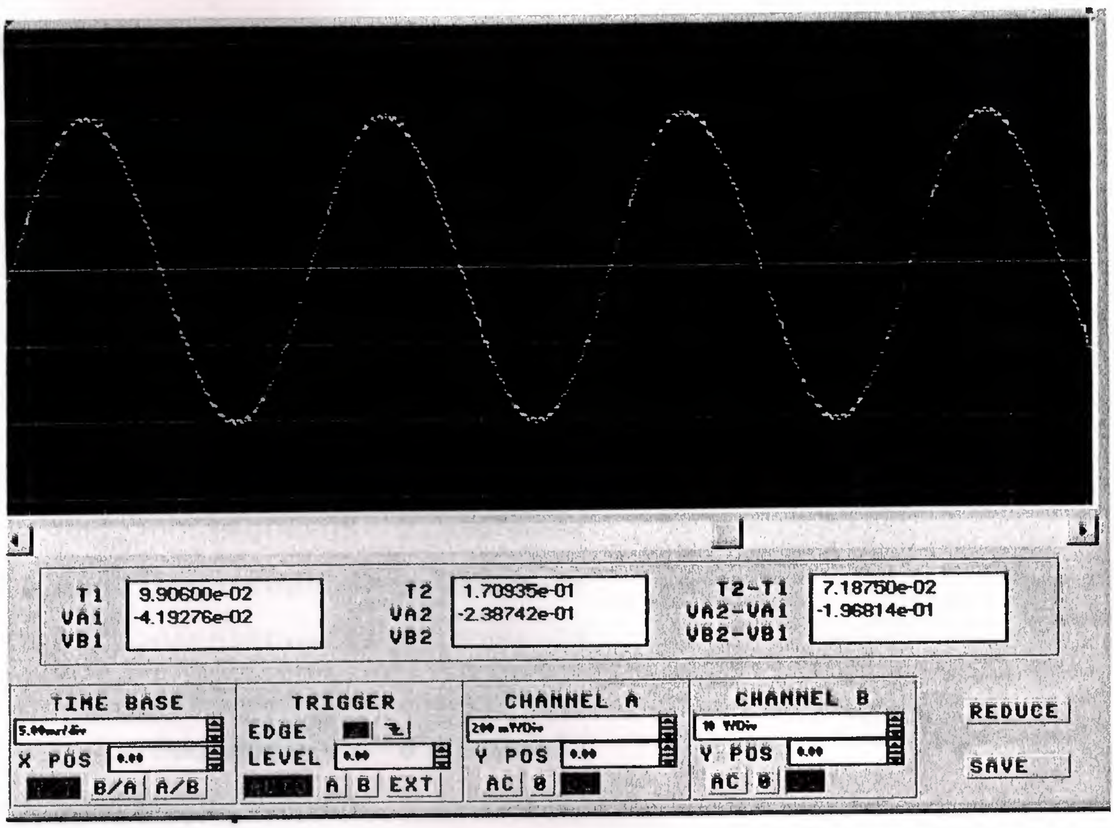

Fig. 3-6-4: The waveform of the low pass output (VL).

Oscilloscope data:

Time base: 0.005 seconds per division

Time offset: 0 seconds

Channel A sensitivity: 0.2 volts per division

Channel A offset: 0 volts

Channel B sensitivity: 10 volts per division

Channel B offset: 0 volts

Channel A connected: yes

Channel B connected: no

Column 1 time (seconds)

Column 2 channel A voltage

Time	Channel A
0.000000000000e+00	0.0000e+00
2.000000000000e-09	1.0000e-05
2.100000000000e-09	1.0000e-05
2.200000000000e-09	1.0000e-05
5.002200000000e-06	2.2477e-05
1.000220000000e-05	5.9895e-05
1.500220000000e-05	1.2226e-04
2.000220000000e-05	2.0956e-04
2.500220000000e-05	3.2180e-04
3.000220000000e-05	4.5898e-04
3.500220000000e-05	6.2111e-04
4.000220000000e-05	8.0817e-04
4.500220000000e-05	1.0202e-03
5.000220000000e-05	1.2571e-03
5.500220000000e-05	1.5190e-03

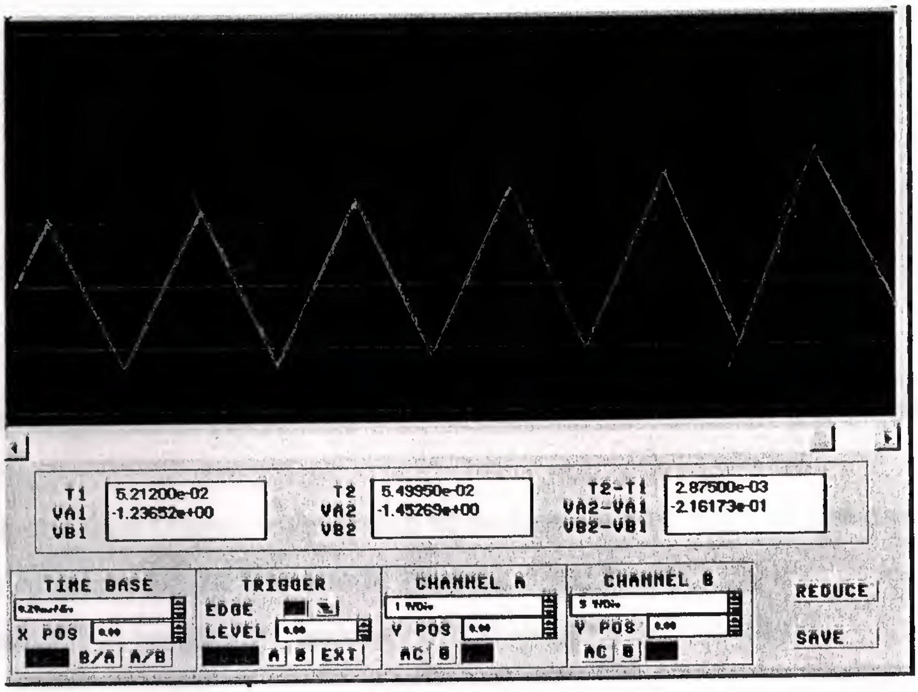

Fig. 3-6-5: The waveform of the band pass output (V_B).

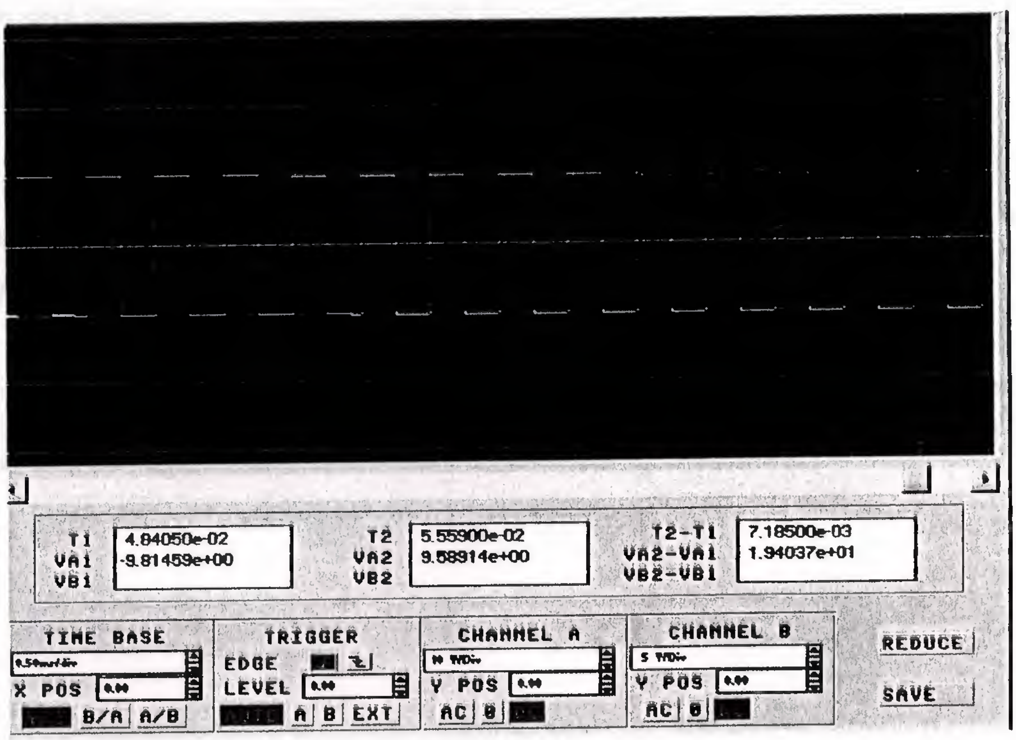

Fig. 3-6-6: The waveform of the anti-resonant output (Vz).

Conclusion

The speech signal has a signal to noise ratio for 3-bit nonuniform Quantizer is 12.1dB,
And for uniform Quantizer the speech signal to noise ratio is 8.4 dB,that mean the signal to noise ratio for nonuniform Quntizer is exceeds by 3.7db than uniform Quantizer, that because the randomly of the speech signal waveform.

The standarad number of Quanization level for speech signal is 256 levels that's needs to a bit rate equal to 64 K. and for speech signal with frequency content between 0-4 KHZ. The sampling frequency Has to be (S=2*4=8.000 sample /second).

To save storage we can usually sample the signal at this minimum rate (Nyquist rate) necessary for the accurate reconstruction at the speech signal .

For spectrum synthesis in part 3.2 the speech synthesis contain from a pulse generator and noise generator and variable frequency resonant , that shown fig. 3-2-2 , and we can control the speech by frequency and the amplitude to obtain asynthsis speech .

For part 3.3 we make calculation of the center frequency and band width , then we note that the speech signal appearnig when the center frequency rang in 0-4 KHZ. That shown in fig.3-4 (variable frequency resoanator).

REFRENCES

1. R.Linggard,'Electronic Synthesis of Speech'1985.
2. John G.proakis/Masoud Salhi,' Communication System Engineering'.
3. L.R. Rbirer/R.W Schafer,'Digital Processing of Speech Signal', 1978.
4. J.N.Holmes,'Speech Synthesis and Recognition', 1988.
5. S.Furuin and M.M Sandhi,' Advanced in Speech Processing', 1990.
6. S.Furui,'Digital Speech Processing, Synthesis and Recognition', 1988.