DOKUZ EYLÜL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

IMPROVING VOICE QUALITY IN VOIP

by
Mehmet Fatih TÜYSÜZ

February, 2007
İZMİR

IMPROVING VOICE QUALITY IN VOIP

A Thesis Submitted to the
Graduate School of Natural and Applied Sciences of Dokuz Eylül University
In Partial Fulfillment of the Requirements for the Degree of Master of Science
in Electrical and Electronics Engineering.

by
Mehmet Fatih TÜYSÜZ

February, 2007
İZMİR

M.Sc THESIS EXAMINATION RESULT FORM

We have read the thesis entitled **"IMPROVING VOICE QUALITY IN VOIP"** completed by **M. Fatih TÜYSÜZ** under supervision of **Asst. Prof. Dr. Zafer Dicle** and we certify that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Asst. Prof. Dr. Zafer DİCLE

Supervisor

Prof. Dr. Mustafa GÜNDÜZALP

(Jury Member)

Assoc. Prof. Dr. Yalçın ÇEBİ

(Jury Member)

Prof.Dr. Cahit HELVACI

Director

Graduate School of Natural and Applied Sciences

ACKNOWLEDGMENTS

I would like to give my sincere thanks to my supervisor, Asst. Prof. Dr. Zafer DİCLE for his guidance, advice and encouragement along the fulfillment of this project.

I want to thank my family for their support and tolarence during this work and all my life, and Ysang ANTIGO who have helped in this thesis.

M. Fatih TÜYSÜZ

IMPROVING VOICE QUALITY IN VOIP

ABSTRACT

Voice quality can articulate lots of things because it depends on people's point of view. First, it is a method of describing and evaluating speech fidelity, intelligibility, and the characteristics of the analog voice signal itself. On the other hand, it can illustrate the accomplishment of the underlying transport instrument. However, voice quality is identified as the qualitative and quantitative amount of the sound and conversation quality of a telephone call.

Due to telephone industry's exchange, existing technologies become valid in different means, and other players become involved. Thus, keeping on the main quality of a telephone call becomes increasingly complicated. Though voice quality has developed over the years to be regularly high and predictable, it is now a significant differentiating issue for new voice over packet (VoP) networks and equipment. As a result, achievement of voice quality in a relatively inexpensive, reliable, and objective way becomes very important.

In this thesis, improving voice quality in Voice over IP and its factors as well as network impairments and their causes in a converged telephony and internet protocol (IP) network, all from the points of view of the network quality, is explained.

Keywords: VOIP, Voice over IP, Echo, Delay, Packet loss, Clarity, Jitter.

İNTERNET ÜZERİNDEN SES İLETİMİNDE SES KALİTESİNİN YÜKSELTİLMESİ

ÖZ

Ses kalitesi, kişinin bakış açısına göre birçok şeyi ifade etmektedir. İlk olarak, aslına uygunluğu, anlaşılabilmeyi ve analog ses sinyali karakteristiklerini tanımlama ve değerlendirme yoludur. Diğer yandan, öncelikli iletim mekanizmasının performansını açıklayabilir. Aslında ses kalitesi, bir telefon görüşmesindeki konuşma kalitesi ve sesin nitel ve niteliksel ölçümleri olarak tanımlanmaktadır.

Telefon endüstrisinin değişimiyle, var olan teknolojiler farklı yollarla uygulanmakta ve yeni oyuncular devreye girmektedir. Böylece, telefon görüşmelerindeki temel ses kalitesinin devamlılığı gittikçe karmaşık bir hal almaya başlamıştır. Ses kalitesinin yıllardır sürekli olarak geliştirilmesine rağmen, şu an yeni ses paket ağları ve gereçleri önemli bir farklılaşma faktörü içerisindedir. Dolayısıyla, ucuz, güvenilir ve objektif bir yol olan ses kalitesi ölçümü çok önemli hale gelmektedir.

Bu tez, internet protokolü ve ağdaki bozulmalar sebebiyle meydana gelebilecek durumları da göz önünde tutarak internet üzerinden ses kalitesinin yükseltilmesini ve faktörlerini açıklar.

Anahtar Sözcükler: İnternet protokolü üzerinden ses iletimi, Echo, Delay, Paket kaybı, Clarity, Jitter.

CONTENTS

CHAPTER ONE
INTRODUCTION

Voice over Internet Protocol (VoIP) is a promising and a fast-growing technology that revolutionized several years ago. It is believed to be the future of the network and the businesses. It offers myriad of benefits to large corporations, medium to small scale businesses and even to individuals alike. It has become an attractive and viable alternative to the traditional Public Switched Telephone Networks (PSTN).

VoIP works through the use of an IP network, enabling the speech signal to be transported in an acceptable way from the sender to the destination. It permits audio and video conversations across an IP based networks which include the internet as well.

This technology uses internet protocol or a packet-switched network that digitizes voice using an audio codec, divides this digitized voice into packets and sends these packets over an IP network to its destination. However, there is no guarantee that all packets routed will travel the same path. Unlike a PSTN call, no dedicated circuit is ever created for a VoIP call.

Four types of communication mode can be made in voice over IP. These are Phone to Phone, Phone to PC, PC to Phone, and PC to PC. Under the first three modes, the voice transmission is carried by both PSTN and IP networks. It requires a VoIP service provider as it as it interconnects PSTN and VoIP networks when a call originates from a PSTN network and arrives at a VoIP network or vice versa.

If a PC connects to some sort of network, it can be used to make calls to anyone who is also connected to that network.

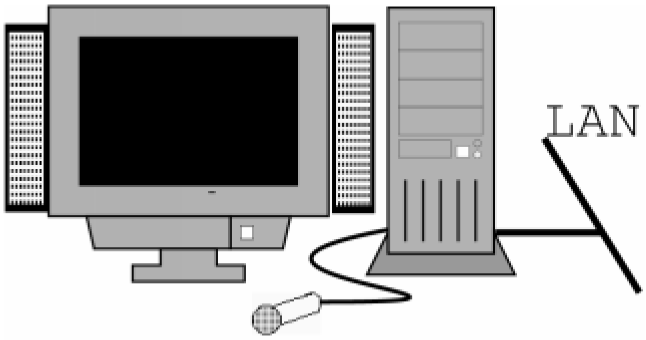


Figure 1.1 PC to LAN configuration

The other situation is a slight variation of the first one. In this situation, a telephone is connected to the PC and used in a similar way as you would when making a normal call. The PC should prepare all the required work to set up the call and to transmit the speech signals. It means that the PC has to be switched on before the call can be made as well. (Liesenborgs, 2000)

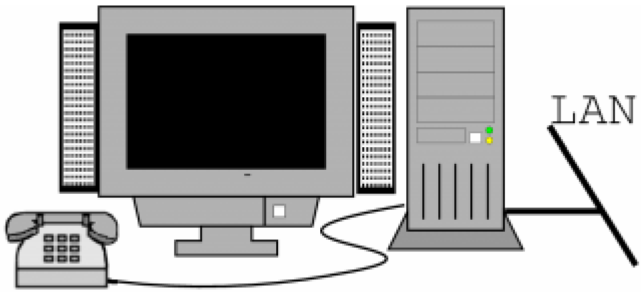


Figure 1.2 Telephone to PC to LAN configuration

Lastly, using of a PC and the necessity of a network could be omitted by the use of a VoIP gateway. It is a particular device that connects the public telephone network with a computer network and performs the essential actions and conversations to make the call possible. Making a call to somebody, you would call the gateway and specify the destination for the call. The call will then be set up and if the other end is available, the conversation can start. This design would be better for people who do not have a PC and It is most likely the easiest to use as well. (Liesenborgs, 2000)

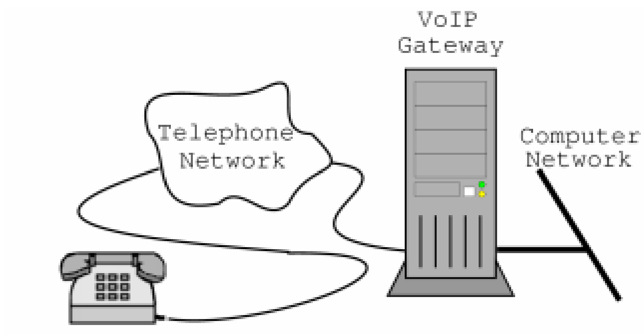


Figure 1.3 Telephone to gateway configuration

1.1 VoIP Protocols

VoIP is the direction finding of voice communications over a network using internet protocol. There are a range of protocols and implementations with a variety of features that are deployed, like for setting up a call, tearing down a call and sending information during a call. Following protocols are the most important VoIP protocols.

Table 1.1 VoIP protocol architecture

Application layer	Audio / Video			
Transport layer	RTP	RTCP	SIP	H.323
	UDP		TCP	
Network layer	IP			
Physical layer	e.g. Ethernet/SDH			

1.1.1 Real-Time Transport Protocol (RTP)

Real-time Transport Protocol (RTP) is an important IETF standard media streaming protocol use to transport real-time data, including audio and video across the network. It handles timing issues as RTP messages contain a sequence number to help detect packet loss, packet duplication or packet reordering.

Table 1.2 RTP header

V=2	P	X	CC	M	PT	Sequence number
Timestamp						
Synchronization source (SSRC) identifier						
Contribution source (CSRC) identifiers						

RTP Control Protocol (RTCP) monitors and suplies data about the RTP flow. Each RTP flow has a corresponding RTCP flow that reports statistics on the call. RTCP is also used for quality of service (QoS) reporting.

The Transmission Control Protocol (TCP) is a good known protocol. It assures reliable and in-order delivery of data from the sender to the receiver.

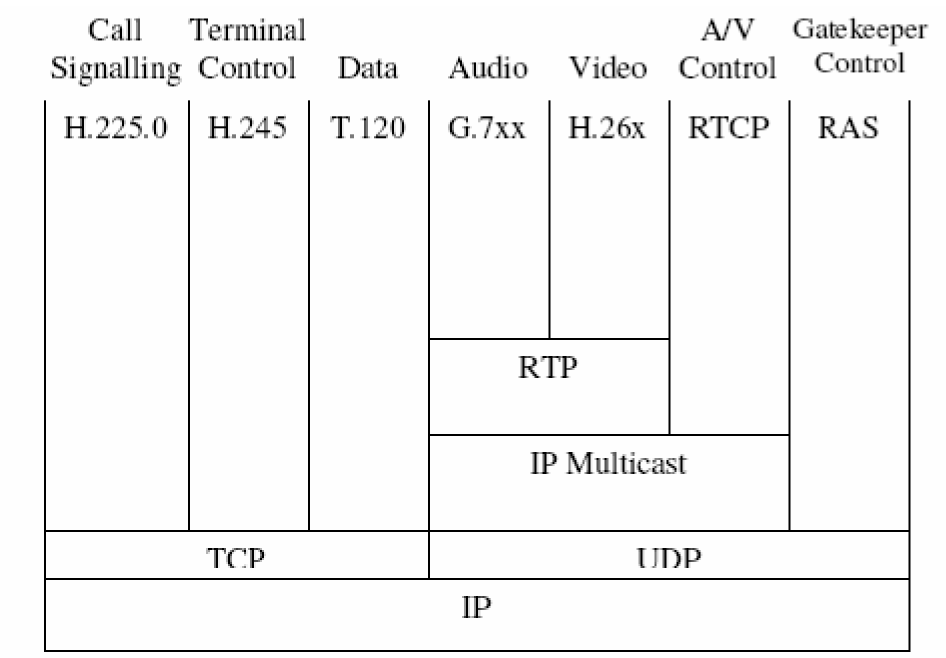
User Datagram Protocol (UDP) is an important protocol of the Internet protocol group as well. Series on networked computers can fling small notes using UDP. These short messages are known as datagrams.

UDP is quicker and has extra capable for many lightweight and time-sensitive intentions. UDP however does not offer dependability and sorting like TCP does. It can be out of use or lost without detect.

1.1.2 H.323

H.323 is the international standard for multimedia communication over packet-switched networks. It is a multimedia conferencing protocol which covers real-time voice, video and data communications. It was originally designed for multimedia in a connectionless environment, such as LAN and has a multipoint voice and video conferencing capabilities. H.323 defines end to end call signaling.(Sinden, 2004)

Table 1.3 The H.323 protocol stack



1.1.3 Session Initiation Protocol (SIP)

SIP is another signaling protocol but simpler than H.323. It is designed to provide a simple way of setting up a call to another user. It handles the setup and the tear down of multimedia sessions between endpoints. Its advantage is that the messages are in text format which means more data are being sent. But it does not mean that the messages are clear and easy to debug.

1.2 VoIP System Structure

Figure 1.5 shows a basic VoIP communication method. There are three parts: the sender, the IP networks and the receiver. At the sender's side, the voice stream from the voice source is first digitized and compressed by the encoder. Then, several coded speech frames are packetized to form the payload part of a packet. The headers are added to the payload and form a packet which is sent to IP networks. The packet may suffer different network impairments (e.g. packet loss, delay and jitter) in IP networks. At the receiver, the packet headers are stripped off and speech frames are extracted from the payload by depacketizer. Playout buffer is used to compensate for network jitter at the cost of further delay (buffer delay) and loss (late arrival loss). The de-jittered speech frames are decoded to recover speech with lost frames concealed (e.g. using interpolation) from previous received speech frames. (Sun, 2004)

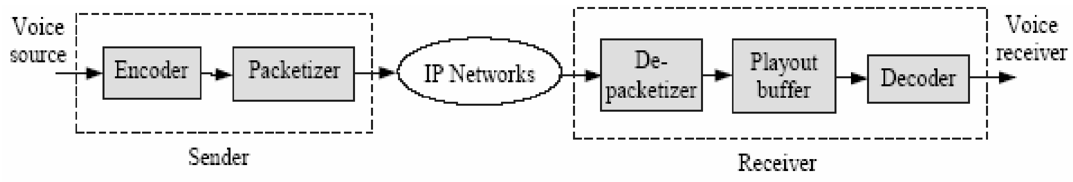


Figure 1.4 Conceptual diagram of a VoIP system

1.2.1 Benefits of VoIP

PSTN networks use a 64 kbps channel for every voice call. It reverses one pipe for every call while in IP networks many calls can use the same pipe simultaneously. It distributes bandwidth between numerous logical links and offloads traffic quantity from existing voice controls. VoIP uses IP networks that have the elasticity to owed bandwidth as required and store the unallocated bandwidth for an additional information. This means that using of network bandwidth in VoIP is more proficient.

Managing and maintaining both voice and data network is not only troublesome but also valuable. Data makes up most important traffic on voice networks. While the voice networks nothing like data networks are not proficient in carrying data due to its restricted and nonflexible bandwidth allowance. The businesses then are forced to preserve both networks. They also have to deal with the trouble of upgrading the voice network equipment such as the Public Branch Exchanges (PBX) telephones. If VoIP is deployed, the voice network will not be required and will go away the enterprises with barely the data network anymore.

VoIP offers a broad range of benefits, thus making it the best alternative to PSTN phone systems. Its benefits include the efficient use of bandwidth, reduction or possible elimination of long distance and phone chargers, convergence of the voice and data networks and advanced features. Some of them will be discussed further below. Another advantage of VoIP is the reduced cost on phone calls or the possible elimination of long distance or phone charges. Service providers of VoIP offer unlimited or fixed number of minutes to make calls which includes long distance calls that is available in monthly flat-rate plans. These plans are much more economical and practical than the traditional charge-by-minute service, thus helping enterprises to reduce their operating costs while maximizing their profits. It is ideal for enterprises and even individuals who make frequent long distance phone calls.(Tse, 2005)

Other services that the traditional phone system offers are also present in VoIP, like speed-dialing, call waiting, busy signaling, caller-ID, etc. However, there are some services that VoIP provides and interoperates and the PSTN don’t. There are video-conferencing, instant messaging, email, click-to-dial and directory services.

1.2.2. Drawbacks of VoIP

Like lots of promising technology there are still a few drawbacks in the system. Because of the direction finding and network latency, the voice information fleeting through an IP network is highly weak to delay and loss. Voice in analog type has to be converted and compressed into digital packets earlier than transmission over an IP network. The density technique insistently minimizes the amount of voice packets thus producing a deteriorated quality of voice. As a result, the quality of voice using VoIP may be not as good as than obtained from PSTN. Its reasons are delay, loss and compression of the information.

While combining voice and data on one platform greatly reduces cost and simplifies management, it also leads to security problems. Voice will be vulnerable to the same attacks as other data that are passing across an IP network. Attacks include interception, modification, spoofing, man-in-the-middle attacks and denial of service. Another drawback of VoIP is the inability to function independently. It relies on properly configured network devices that are dependent to a stable electrical power supply. Therefore during power outages there is no VoIP phone service available, unlike the traditional phone that is kept in service. (Tse, 2005)

CHAPTER TWO
IMPROVING VOICE QUALITY

Voice quality is very important for communication. It is a mission-critical application and is based on speech transmission. With the emerging popularity of VoIP in the past few years, many believe that this technology will continue to improve both in the carrier and enterprise sectors. It is also claimed that VoIP will develop from being a substitution for PSTN into supllying an entirely converged services to residences and companys.

Voice quality is the foremost issue in VoIP. It is one of the hardest services to provide in packet-switched networks. The PSTN was built to provide an optimal service for time sensitive voice applications, with low delay, low jitter and constant but low bandwidth. On the other hand, IP networks have been built to support non-real time data applications such as email or file transfer. The applications are characterized by bursty traffic, with occasional peaks in demand for high bandwidth, but are not sensitive to delays. In a conversation, humans have little tolerance to delays, jitters, echoes and noises.(Kauffman, 2006)

As soon as a voice is transformed from analog electrical signals to a digital one, it is being compressed resulting to some lost components. In order to provide an improved voice service using this platform, some components of voice quality must be maintained. The method used to measure its quality and the quality of service (QoS) tools that can be implemented in a network must be discovered as well.

2.1 Improving Voice Quality

Supervisors appear dispute in distributing voice traffic properly because of the natural features of a converged voice and data IP network. Challenges will be depicted and even the explanations for shuning and overcoming them when intending a VoIP network for the most advantageous voice quality.

2.1.1 Network Quality

A VoIP system with poor network quality reduces its performance. In VoIP applications, delay, and packet loss are important network impairments that affect the voice quality. Jitter problems can be solved by using a playout buffer at the receiving end but it can also increase delay and additional packet loss. There are a few impairment types (logical and physical) in the internet protocol network that cause delay, jitter and packet loss. (Kauffman, 2006)

There are some functions that can be used on network to outcome in delay, jitter and packet loss. Some of these components are:

- ✓ Network protocols - routing protocols, traffic control protocols
- ✓ Router process
- ✓ Bandwidth of the connections
- ✓ Network dependability

Dependability of network is an imperative part in VoIP that launchs delay and packet loss, principally in the spinal column of IP networks. Two key problems can straightly impact the network dependability and these are routing reconfiguration and connection failures.

- ✓ Connection failure:

A connection breakdown characteristically appears as a period of repeated packet loss that can lost for many seconds followed by a change in delay after the link is re-established. It can be caused by equipment problems, cut or unplugged cables, a configuration change in a transport network or the denial of service attack. Link failure will result in significant gaps in received speech.(Kauffman, 2006)

✓ Routing reconfiguration:

When a connection turns off it is common for a routing protocol to need around 5 seconds to congregate to a new configuration and around 15 seconds when a link goes up. for the duration of this reconfiguration time, forwarding may be interrupted and voice packets may be lost. (Kauffman, 2006)

2.1.2 Factors that Affect Voice Quality

There have been a lot of developments in Voice over IP. But since the speakers and listeners are already used for years to the faultless quality of landline phone, they don’t have good ideas about what voice quality is. In designing a VoIP network, it is significant to think all the causes that will influence voice quality. The most significant obsessions that influence quality of voice transmited over IP and explanations will be clarified to develop the voice quality.

Due to the scenery of IP networking, definite communication troubles most likely occur when voice packets are sent via IP. Circumstances present in the network can establish troubles like echo, jitter, or delay. These troubles have to be solved with QoS tools.

clearness of digital audio signal is significant for suitable phone call. Listener should be able to distinguish the speaker's characteristic and be aware of what are they talking about. These features can influence clearness:

- Fidelity, The degree to which a system or a portion of it accurately reproduces at its output, the necessary characteristics of the signal impressed upon its input or the result of a prescribed operation on the signal impressed upon its input. The total bandwidth of the spoken voice is almost always limited by the bandwidth of the transmission medium. On the average, the human speech requires a bandwidth of 100 to 10,000 Hz, although 90 percent of the speech intelligence is contained between 100 and 3000 Hz. (Wallace. 2006)

- Jitter is a disparity of package delay where delays actually impact the quality of the conversation. It occurs when voice packets are sent and received with timing variations. It causes various spaces in dialogues that are undesirable and annoying to the listener.

- A jitter buffer stores arriving packets for the time being in order to minimize delay variation. It assigns small buffer to receive the packets and gives it to the receiver with small delay. If packets arrive too late then they are discarded which leads to call quality degradation. Usually in IP telephones (hardware and software) buffer lengths can be modified. If jitter buffer is increased it turns out in less packet loss but more delays. A reduction turns out in less delay but more packet loss.

- Echoes happen due to a incompatible hybrid (2 to 4 wire convetors) on the analog part of a telephony connection. It is a consequence of electrical impedance mismatches in the communication path. One more source of echo is acoustic response from speaker to microphone of a phone receiver. It is forever there in traditional telephony networks, but at a point that cannot be detected by the human ear.(Wallace. 2006)

- Packets are lost due to some reasons. It takes place when a large amount of traffic hits the network and it causes to drop packets where in 20 ms of audio is lost. It usually manifests itself as dropped conversation or “tinny” sounds. To stay away from the packet loss troubles the most useful method is to not send silences (particularly in squat speed networks or with blockage). dialogues have a lot of silence instants.

- Delay or latency is the time between a spoken voice and the arrival of the digitized one transported at the far end. Physical distance, the number of router hops, encryption and voice/data conversion all impact latency. It is not only a problem in VoIP but also in telecommunication networks.

- Bandwidth determines connection speed. It plays an important role in delivering a good voice quality. For a dial-up connection, not much should be expected. A broadband connection will work right as lot as it is not spotty and not shared with too many other communication applications. It implies that the greater the connection speed, the better the voice quality you can get.

- Hardware equipment used can significantly influence the quality of a dialogue in voice over IP. Usually reduced quality equipment are the cheapest, but not for all time.

- Codec is the software used to compress voice packets that are being transported over IP network so that the load transmitted is lighter. Some codecs are better than others and each codec is designed for a definite role. If a codec will be used for a statement requirements, it should be preferred in any speech codecs.

Background noises come in many different shapes and sizes that is heard from the far-end connection. definite bandwidth - saving equipments such as voice activity detection (VAD) can reduce this backdrop noises in total. When this knowledge is applied, the speaker audio trail is unlock to the listener, while the listener audio trail is closed to the speaker. The result of VAD is frequently that speakers consider that the link is broken, since they hear nothing from the other end. (Wallace, 2006)

2.2 Quality Metrics

Voice quality must be computable in order to classify. There are three quality metrics consist of the Mean Opinion Score (MOS), the Perceptual Speech Quality Measurement (PSQM), and the Perceptual Evaluation of Speech Quality (PESQ).

2.2.1 Mean Opinion Score (MOS)

Mean opinion score is a voice call quality metric. It is the most famous measure of voice quality. It is a scoring system and subjective method of quality assessment. It efforts with the two test process, dialogue opinion test and listening opinion test. The quality of voice communication structure is judged through carrying on a conversation or by listening to speech samples. They grade the voice quality using the following scale: (Spirent, 2001)

5– Excellent, 4–Good, 3 – Fair,2 –Poor, 1 – Bad

MOS was formerly proposed to evaluate the quality of various coding standards. The following is a outline of the MOS for various coding algorithms.

Table 2.1 MOS scores for different codecs

Compression Method	Bit Rate (Kbps)	Sample Size (ms)	MOS Score
G.711 PCM	64	0.125	4.1
G.726 ADPCM	32	0.125	3.85
G.728 Low Delay Code Excited Linear Predictive (LD-CELP)	15	0.625	3.61
G.729 Conjugate Structure Algebraic Code Excited Linear Predictive (CS-ACELP)	8	10	3.92
G.729a CS-ACELP	8	10	3.7
G.723.1 MP-MLQ	6.3	30	3.9
G.723.1 ACELP	5.3	30	3.65
iLBC Freeware	15.2	20	3.9
	13.3	30	

Helpers take note the voice samples and range them from 1 to 5, where 1 is the worst and 5 is the best. The test scores are averaged to a combination score. The test results are subjective since they are based on the beliefs of the listeners.

2.2.2 Perceptual Speech Quality Measure (PSQM)

The automated process of measuring speech quality is called Perceptual Speech Quality Measure. It usually be located with IP call managing systems. It exactly works out the dissimilarities between the input and output signals.

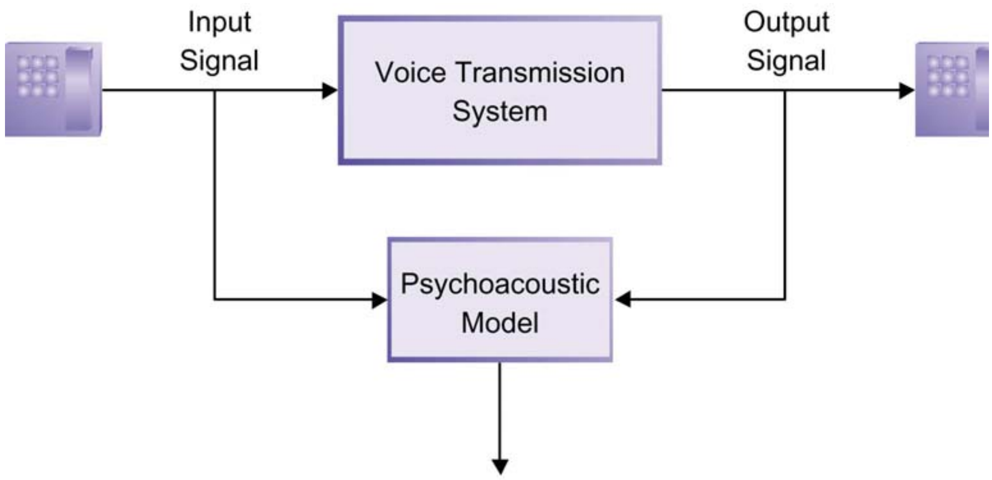


Figure 2.1 PSQM

At this technique, the PSQM score will be zero if the input and output matches. The bigger differences, the higher the score will be up to the highest of 6.5. The stress of PSQM is on the differences that will influence person observation of speech quality, unlike other conventional measurements such as signal to noise ration (SNR). Apparatus and software that can assess PSQM is obtainable through third-party vendors. The PSQM measurement is made by comparing the original transmitted communication to the resulting speech at the far end of the transmission canal. This system is made to be deployed as in-service components. The PSQM measurements are made during real conversation on the network. Unlike the subjective listening test, this automated testing algorithm is over 90 percent accurate. Scoring is based on a scale from 0 to 6.5, where 0 is the best and 6.5 is the worst. PSQM does not take into account the jitter or delay problems that are experienced in packet-switched voice system since it was originally designed for circuit-switched voice.(Spirent, 2001)

2.2.3 PESQ

MOS and PSQM were intended earlier than the appearance of VoIP knowledges. It is not sufficient for voice over IP network quantity. It does not compute representative issues in VoIP such as delay and jitter. Since MOS assessor has no concept of two-way conversation and only listens to audio quality, it is feasible to attain a score of 3.8 in MOS on a VoIP network when one way delay exceeds 500 ms. The one way delay is not appraised.

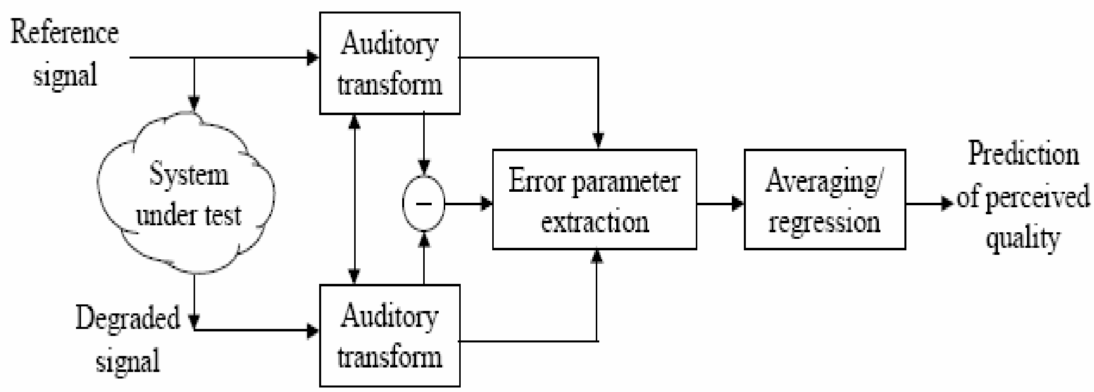


Figure 2.2 PESQ

PESQ has evolved into ITU standard P.862 which is considered as the current standard for voice quality measurement. Its function as illustrated in figure 2.2 was originally developed by British Telecom, Psytechnics, and KPN Research of the Netherlands. The PESQ can take into account codec errors, filtering errors, jitter and delay problems that are normal in a VoIP network. It combines the best of the PSQM method along with the method called Perceptual Analysis Measurement System (PAMS). Its scores range from 1 (worst) to 4.5 (best), with 3.8 considered "toll quality" (that is, acceptable quality in a traditional telephony network). It is meant to measure only one aspect in voice quailty. PESQ scores however does not reflect the effects of a two-way communication such as loudness, loss, delay, echo and sidetone. (Wallace, 2006)

CHAPTER THREE
CODECS

Voice over IP needs compression and fortunately voice information offers the possibility of large compression ratios. Regarding to Voice over IP, a codec is an algorithm in order to encode and decode the voice conversation. As it is heard voice is analogue and it needs to be converted (or encoded) to a digital format suitable for transmission via internet. Firstly, it is necessary to decode it again so the other person should understand what is it said about. There exist different encoding and decoding ways, many of them operate compression to lessen the necessary bandwidth of the conversation. An important thing to keep in mind with Voice over IP, is that encoding, mostly when profound compression is used, takes some period of time, which inserts a delay to the conversation. So, the milestone is a codec which not only preserves good quality with compression, also it can encode and decode in a minimal period of time as well.

There are a few kind standard speech codecs used on Voice over IP. PCM, ADPCM, CELP are the most importants of them.

3.1 Standard Speech Codecs

There are a few speech codecs are described in the below;

- 64 kbits/s PCM Codecs
- The 32 kbits/s G721 ADPCM Codec
- The 16 kbits/s G728 Low Delay CELP Codec

3.1.1 PCM Codecs

The simplest type of waveform codecs is Pulse Code Modulation (PCM) codecs. For speech coding it was created that with nonlinear quantization 8 bits for each model was enough for speech quality which is almost impossible to differentiate from the original. It provides a 64 kbits/s, and two such nonlinear PCM codecs were regulated in the 1960s. U-law coding alghorithm is used in America however, in Europe, A-law coding alhorithm is used. Since they are easy to process, have great quality and low delay both these codecs are still broadly used at present. For instance, the .au audio files that are frequently used to transmit sounds over the network.(Woodard, 1995)

3.1.2 ADPCM Codecs

Adaptive Differential Pulse Code Modulation (ADPCM) codecs are waveform codecs. They quantize the diversity between the speech signal and a forecast that has been made of the speech signal like PCM codecs, in place of quantizing the speech signal directly. The variation between the predicted and real speech samples will include a lesser variation than the real speech samples if the prediction is accurate and will be correctly quantized with fewer bits than would be required to quantize the original speech samples. At the decoder in order to give the renovated speech signal, the quantized difference signal is added to the predicted signal. The performance of the codec is assisted by using adaptive prediction and quantization, in order that the predictor and difference quantizer adapt to the varying characteristices of the speech being coded. In the middle 1980s the CCITT standardised a 32 kbits/s ADPCM, in other words G721, which gave reconstructed speech almost comparable with the 64 kbits/s PCM codecs. Afterwards in recommendations G726 and G727 codecs functioning at 40,32,24 and 16 kbits/s were standardised.(Woodard, 1995)

3.1.3 The G728 Low Delay CELP Codec

Lesser quality of waveform codecs and around of 16 kbits/s bit rates falls fastly. Therefore at these rates hybrid codecs, particularly CELP codecs and their derivatives, have a tendency to be used. Though due to the forward adaptive determination of the short term filter coefficients used in most of these codecs, they tend to have high delays. The delay of a speech codec is described as the time from when a speech sample arrives at the input of its encoder to when the corresponding sample is produced at the output of its decoder, guessing the bit stream from the encoder is fed through the decoder. This delay will be of the order of 50 to 100 ms, and such a high delay can create problems for a typical hybrid speech codec. So in 1988 the CCITT released a set of requirements for a new 16 kbits/s standard, the main necessities being that the codec should have speech quality comparable to the G721 32 kbits/s ADPCM codec in both error free conditions and over noisy channels, and should have a delay of less than 5ms and idealy less than 2ms. (Woodard, 1995)

3.2 Compression Standards

To provide a feedback between feasible applications, it is significant that standards are recognized. The most broadly familiar standards in the Voice over IP domain, are the G. standards of the ITU-T. Other well known standards are the ETSI GSM standards. There is a table of some standards:

Table 3.1 G.7xx Codecs

Standard	Description	Bit rate	MOS
G.711	Pulse Code Modulation using eight bits per sample, sampling at 8000 Hz	64 kbps	4.3
G.723.1	Dual rate speech coder designed with low bit rate video telephony in mind [41]. The G.723.1 coder needs a 7.5 ms lookahead and used one of these coding schemes: • Multipulse Maximum Likelihood Quantisation (MP-MLQ) • Algebraic CELP (ACELP)	6.3 and 5.3 kbps respectively	4.1
G.726	Coder using ADPCM. Contains obsolete standards G.721 and G.723	16,24,32 and 40 kbps	2-4.3
G.727	Five, four, three and two bits per sample embedded ADPCM. The encoding allows bit reductions at any point in the network without the need for coordination between sender and receiver [10].	16,24,32 and 40 kbps	2-4.3
G.728	Low Delay CELP (LD-CELP)	16 kbps	4.1
G.729	Conjugate Structure ACELP (CS-ACELP) • Annex A: Reduced complexity algorithm • Annex D: Low rate extension • Annex E: High rate extension These coders need a 5 ms lookahead.	8 kbps (CS-ACELP), 8 kbps (Annex A), 6.4 kbps (Annex D) and 11.8 kbps (Annex E)	4.1 (CS-ACELP) and 3.7 (Annex A)
GSM	Full rate speech transcoding using Regular Pulse	13 kbps	3.71

06.10	Excitation-Long Term Prediction (RPE-LTP)		
GSM 06.20	Half rate speech transcoding using Vector Sum Excited Linear Prediction (VSELP)	5.6 kbps	3.85
GSM 06.60	Enhanced full rate speech transcoding using ACELP	12.2 kbps	4.43

Formerly, Mean Opinion Score information about some coders could not locate. The MOS are quite individual and it is possibly because of the MOS values frequently differ according to different sources. Occasionally these dissimilarity are even quite large.

Consequently, for telephone quality statement using digitised speech, a bandwidth of 64 kbps is required if the speech data is absent uncompressed. However, speech data can often be significantly compressed and this can drastically decrease the amount of needed bandwidth.

Some compression techniques don't work out the nature of data. Some techniques recommend some compression, but generally they do not effect in high density ratios. Though, they can be used to more decrease the required amount of storage when a different compression technique has already compressed information of the voice. Waveform supposes that the data is exist with an audio signal, but in general they don't exploit the fact that the signal holds only speech data. They attempt to model the waveform as intimately as possible. The results have good speech quality at high data rates. Certain amount of delay into the communication introduced by compressing and decompressing speech data. The amount of lookahead that a compression system requires is most likely the most significant delay factor because computers are becoming ever faster and specialised hardware is becoming accessible. It is significant that standards are recognized in order to supply interoperability between diverse applications. (Liesenborgs, 2000)

3.3 Affect of Compression Algorithm on Voice Quality

Speech class is cooperated to the bit rate of the signal. Commonly, lower bit rates means lower perceived quality. Table 3.1 demonstrates the connection between speech compression bit rate and sound quality.

Table 3.2 Relationship between bit rate and speech quality

Bit Rate (k bps)	**Speech Quality**
64 (or greater)	Broadcast
64 to 12	Toll
12 to 6	Communications
Below 6	Synthetic

Maximum quality of speech is called as broadcast quality. Broadcast quality is like a voice quality on a cd. Toll quality is the quality accomplished in usual telephone systems. Connection quality is illustrated as being comprehensible but noticeably lower quality than toll due to distortion. Synthetic quality speech is intelligible, but sound is unusual. An conduct test was carried out to measure the perceived sound quality of audio codecs frequently used by Internet Telephony applications. The test isolated the effect of compression algorithm on voice quality before other network induced impaiments were introduced into the system. The quality of the signals were charged by a group of 16 applicants. The test signals and consequences of the study are shown in Table 3.2 (Sunstrom, 1999)

Table 3.3 MOS Rating of Internet Telephony Compression Algorithms

Test Signal	**Voice**	**Audio Codec**	**MOS**
1	Male	PCM	4.0714
2	Female	PCM	4.0
3	Male	True Speech	3.9286
4	Female	True Speech	3.7857
5	Male	GSM	4.0
6	Female	GSM	4.0
7	Male	A-law	3.7143
8	Female	A-law	4.4286
9	Male	u-law	4.4286
10	Female	u-law	4.0714

The results of the MOS survey confirm that the algorithms used in VoIP systems are able to accomplish toll quality voice. Quality restrictions in Voice over IP schemes are because of reasons of network induced objects like delay and packet loss. (Sunstrom, 1999)

CHAPTER FOUR
PACKET LOSS

In the PSTN, a call is allocated with a physical link between endpoints, and the circuit stays dedicated to that channel for the duration of the call. If it is compared to packet networks, packet networks break voice, fax, and data into small samples or packets of information. Each packet has a header that identifies where the packet is going and provides information for reassembly when the packet arrives at the destination. Packets travel independently and they are interspersed with packets from other network traffic along the way. Travel time through the network varies for individual packets. Unless the network is precisely matched to the peak traffic load, packets sometimes fail to arrive at the destination. These lost packets create gaps in voice communications, which can result in clicks, muting, or unintelligible speech. In transmitting data, the remedy for packet loss is to resend the missing packets, but this solution doesn’t work for time-sensitive voice conversations. Generally, there are two ways to lose packets. They can be lost at network nodes because of an over-flow in the buffer or because a congested router deliberately discards them to reduce congestion. These packets are strictly gone, and will never reach at the destination.(Nortel, 2001)

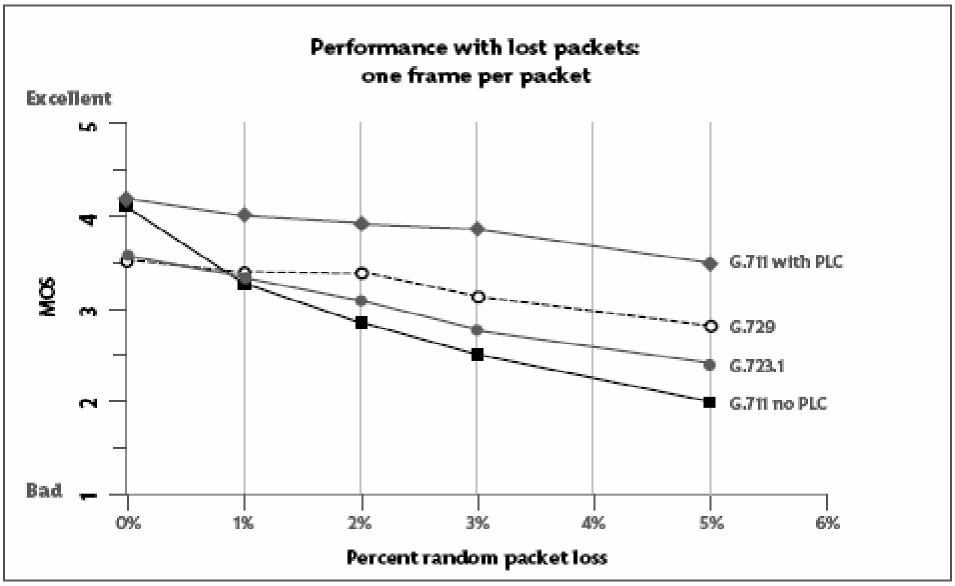


Figure 4.1 Packet loss effects for three common speech codecs

After one or more models journey across a computer networking becomes unsuccessful in order to reach their target packet loss occurs. There are a lot of factors can force to the packet loss. Such as signal degradation over the network medium, oversaturated network links, corrupted packets rejected in transit or faulty networking hardware.

Some network transport protocols such as TCP provide for reliable delivery of packets. At packet loss, receiver requests for retransmission or the sender automatically resends any segments that have not been acknowledged. Although TCP can recover from packet loss, retransmitting missing packets causes the throughput of the connection to decrease. This drop in throughput is due to the sliding window protocols used for acknowledgement of received packets. In some protocols, if a transmitted packet is lost, it will be resent along with every packet that had been sent after it. This retransmission causes the overall throughput of the connection to drop.(Wikipedia, 2007)

User Datagram Protocol do not supply recovery for lost packets because it is devised to handle this type of packet loss.

4.1 Loss Distribution

In general, packet loss distribution in Internet Protocol networks are called "bursty" but there is fewer assurance in relation to the use of specific loss models, and indeed, some misunderstanding related to a few usually used types, for instance the Gilbert Model.

4.1.1 Historical background

Some simple works regarding loss or error modeling found out in the 1960's in relation to the delivery of bit errors on telephone channels. One advance used was a Markov or multi-state model. Gilbert appears to be the first to describe a burst error model of this type, later extended by Elliott and Cain and Simpson. Blank and Trafton produced higher state Markov models to represent error distributions. Another approach was to identify the statistical distribution of gaps. Mertz used hyperbolic distributions and Berger and Mandelbrot used Pareto distributions to model inter-error gaps. Lewis and Cox found that in measured error distributions there was strong positive correlation between adjacent gaps. Packet loss modelling in IP networks seems to have followed a similar course, although the root cause of loss (typically congestion) may be different to that of bit errors (typically circuit noise or jitter). (Voiptroubleshooter, 2004)

Gilbert Model lossy state matchs to a "loss" state, i.e. that the possibility of packet loss in state 1 is 1 but this is incorrect (it would be more suitable to explain this as a 2 state Markov model). Markov model is a general multi state model in which a system switches between states i and j with some transition probability p(i, j). A 2-state Markov model has some merit in that it is able to capture very short term dependencies between lost packets, i.e. consecutive losses. These are generally very short duration events (say 1-3 packets in length) but occasional link failures can result in very long loss sequences extending to tens of seconds. By combining the 2-state model with a Gilbert-Elliott model it is probable to confine together very small period following loss events and longer lesser density events. (Voiptroubleshooter, 2004)

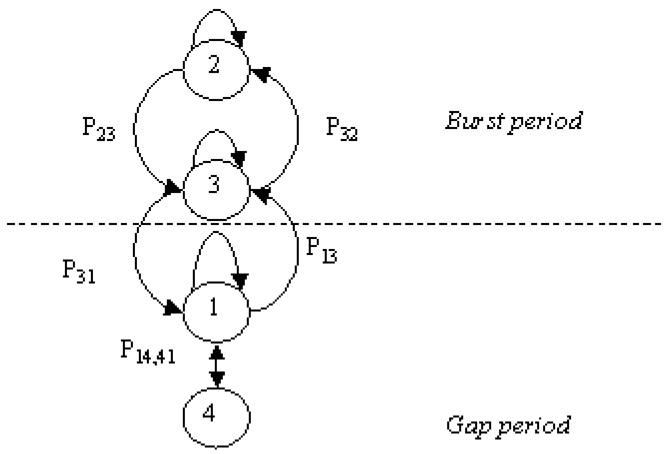


Figure 4.2 State Markov Model

4.2 Packet Loss Measurement

The IETF sketchs a attitude for measuring packet Ioss. This method is summarized below.

1. Synchronize the source and destination clocks
2. Send a packet fiom source to destination that contains the departure time
3. Timestamp the packet upon its arrival at the destination.
4. Subtract the departure time from the arriva1 time.
5. If the packet arrives within the allowed threshotd. count the packet as received.
6. If the packet fails to arrive within the allowed threshold. count the packet as discarded. (Sundstrom, 1999)

Attitude presented by the IETF for measuring packet loss is very related to that presented for measuring end to end delay. Actually, the only disparity is what is reported as the rate of the metric. At the tests, necessity is for a packet to reach at the target within an suitable entrance. Entrance must be similar as that which was classified for packet delay: 150 - 400 ms. If the packet does not reach inside the acceptable entrance, then it is too late to be used.

4.2.1 Packet Loss In the Intemet

The typical packet loss rates for every continent are summarized in Table 4.1.

Table 4.1 Approximate Average Packet Loss in the Internet

Continent	Average Packet Loss Rate (%)
North America	2
Asia	7
Australia	5.5
South america	5.5

4.2.2 Effects of Packet Loss On Voice Quality

Packet loss causes some loss of information on a voice conversation. The quantity of packet loss endured by a Voice over IP application is relative to the quality of the communication.

An exploding sound consequences when a packet is crashed from a flow of speech packets. The object is a result of the discontinuity in amplitude between one section of speech and the missing section.

4.3 lmproving the Quality of Speech

It is clear that packet loss decompose the quality of voice communication. The information included in the missing packet must be put back. Packet recovery methods are a existing area of work. The most straightforward way of replacing the information lost in the missing packet is to replace the packet with noise. This has been shown to be an improvement in quality compared to structures that simply play out the noiseless interval. an additional technique of recovering the information is to play again the last properly received packet in the place of the missing packet. This method has been shown to be an development over replacing the packet with sound.

This method can be successful if packet loss is occasional and happens in non-continuous blocks. While this method fills in the missing period. it does not restore the missing data. (Sundstrom, 1999)

A different recovery method is to convey redundant data about the n^{th} packet along with the n+1 packet. Some differences of this method have been proposed. They sort from carrying the whole preceding packet along with the next packet to carrying only properties of the n^{th} packet along with the n+1 packet. The redundant properties carried by the n+1 packet are frequently produced by a vocoder. The benefit of this method is improved statement quality even under high degrees of packet loss. However, end-to-end delay enlarges by the time needed to encode the redundant data which in turn affects the VoIP performance. Besides, the application bandwidth increases, which may donate to blockage and ultimately packet loss. Packet recovery techniques are presently an open area of research in VoIP.(Sundstrom, 1999)

4.3.1 Packet Loss Recovery

Fault managing system is needed if the amount of missing audio packets is higher than that beared by the listener at the receiver end. Characteristic systems drop in one of two modules. Automatic Repeat Request (ARQ) systems are closed ring systems stand for the reconduction of the packets that were not expected at the receiver end. Forward Error Correction (FEC) systems are open ring systems stand for the conduction of superfluous data along with the innovative data so that the missing innovative data can be improved from the superfluous data. ARQ systems are not satisfactory for live audio functions like audio conferencing over the Internet since they noticeably enlarge end to end latency. Besides, they are not well-matched to multicast surroundings.

FEC is an smart option to ARQ to offer dependability without growing latency. It is principally significant for purposes with real-time restrictions over high speed networks. Nevertheless, the possiblity of FEC systems to improve losses depends significantly on the individuality of the packet loss method in the network. Obviously, FEC systems are more successful when missing packets are isolated during the flow of packets sent from a sender to a receiver end.

The easiest method to insert idleness to an audio packet is to attach no idleness at all. Certainly, it is probable to pick up at the receiver end from packet losses without any superfluous data. For instance, a lost packet can be put back by silence or by noise. A better method is to rebuild it by reproducing the earlier packet.

4.3.2 Packet Loss Concealment

Packet Loss Concealment is a well known method that used to cover the influences of missing or superfluous packets. For instance, a G.711 packet was lost, the Voice over IP mechanism migt select to basically play again the missing packet to mask the fact that a packet was lost, rather than tolerating the user to pay attention to stillness. While that is an exceptionally uncomplicated method, there are more superior algorithms to supply packet loss concealment, with some codecs (e.g., G.729) containing PLC as an fundamental element of the intend. PLC is usually useful only for little amounts of repeated missing packets. For instance, a whole of 20 - 30 milliseconds dialogue and for short packet missing charges.

Packet loss migth seem bursty in real - with times of some seconds throughout which packet loss may be 20-30 percent. The standard packet loss charge for a conversation migth be short. Nevertheless, this era of high loss charge may cause perceptible dreadful conditions in conversation quality.

PLC techniques usually engage either playing again the missing packet received or several more complicated techniques that uses earlier speech samples to produce speech. Easy repeat techniques be likely "robotic" sounding dialogue when several repeated packets are lost. More complicated techniques may supply logical quality at 20% packet loss charges. Nevertheless, it can devour DSP bandwidth and therefore it causes lessen the amount of channel that can be holded in.

CHAPTER FIVE
DELAY

Sometimes voice packets take more time than thought to reach their target and it causes delay on Voice over IP networks. It causes some disruption in the voice quality as well. However, if delay is dealt with truthfully, its effects can be minimized.

When voice packets are transmited over a network towards a destination machine/phone, some of them might be delayed. Reliability features in the voice quality mechanism sees to it that a conversation is not deadlocked waiting for a packet that went to have a walk somewhere in the green. In fact, there are many factors affecting the journey of packets from source to destination, and one of them is the underlying network. When a packet is delayed, you will hear the voice later than you should. If the delay is not big and is constant, your conversation can be acceptable. Unfortunately, the delay is not always constant, and varies depending on some technical factors. This variation in delay is called jitter, which causes damage to voice quality. Delay causes echo in VoIP calls.(About, 2000)

5.1 Delay Limits

The International Telecommunication Union (ITU) takes into account network delay for Voice over IP applications in commendation G.114. This commendation describe three bands of one way delay as shown in Table 5.1.

Table 5.1 Delay Specifications

Range in Milliseconds	**Description**
0-150	Acceptable for most user applications.
150-400	Acceptable provided that administrators are aware of the transmission time and the impact it has on the transmission quality of user applications.
Above 400	Unacceptable for general network planning purposes. However, it is recognized that in some exceptional cases this limit is exceeded.

These commendations are familiarized for national telecom administrations. These are more strict than when usually applied in secretive voice networks. When the position and production requires of end users are well known to the network designer, more delay can confirm satisfactory. For classified networks 200 ms of delay is a practical aim and 250 ms a limit. All networks require to be engineered such that the maximum expected voice connection delay is known and minimized. (Cisco, 2000)

5.2 Sources of Delay

In a voice conversation, a call participant is aware of the time taken for the remote user to react. This time includes the round trip delay experienced by the voice signal and the reaction time of the remote user. The user is not aware of any asymetry that may be present in the time taken by the outgoing versus incoming voice signal. Various delays can be measured or specified. Round trip delay refers to the total delay for the sending and receiving direction combined. One-way delay refers to the delay in either the sending or receiving direction. Symmetric one-way delay refers to the delay in the sending or receiving direction with the assumption that they are equal (Telchemy, 2006)

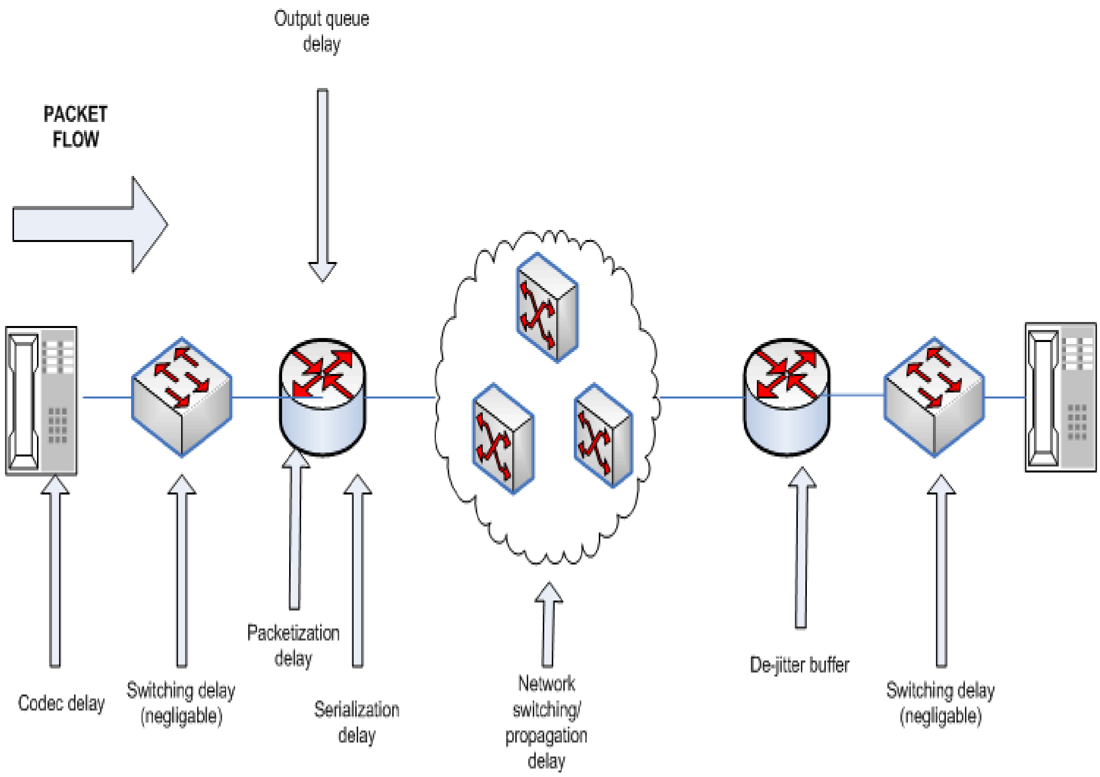


Figure 5.1 Delay Sources

It is clear in the figure that, delay occurs from some reasons. Codec delay, packetization delay, serialization delay, propagation delay and network switching delay are some types of delay sorts. It means that they cannot be reduced with any of the Quality of Service gears.

5.2.1 Coder Delay

Coder or processing delay is the needed time period to compress a block of PCM samples by the digital signal processor (DSP). It is called processing delay (χ_n) as well. This delay changes with the sound coder used and processor speed. For instance, algebraic code excited linear prediction (ACELP) algorithms analyze a 10 ms block of PCM samples, and then compress them. The compression time for a Conjugate Structure Algebraic Code Excited Linear Prediction (CS-ACELP) process ranges from 2.5 ms to 10 ms based on the loading of the DSP processor. If the DSP is fully loaded with four voice channels, the Coder delay is 10 ms. If the DSP is loaded with only one voice channel the Coder delay is 2.5 ms. For design purposes use the worst case time of 10 ms. Decompression time is roughly ten percent of the compression time for each block. However, the decompression time is proportional to the number of samples per frame because of the presence of multiple samples. Consequently, the worst case decompression time for a frame with three samples is 3 x 1 ms or 3 ms. Usually, two or three blocks of compressed G.729 output are put in one frame while one sample of compressed G.723.1 output is sent in a single frame. (Cisco, 2000)

Best and worst case coder delays are shown in Table 5.2.

Table 5.2 Best and Worst Case Processing Delay

Coder	**Rate**	**Required Sample Block**	**Best Case Coder Delay**	**Worst Case Coder Delay**
ADPCM, G.726	32 Kbps	10 ms	2.5 ms	10 ms
CS-ACELP, G.729A	8.0 Kbps	10 ms	2.5 ms	10 ms

MP-MLQ, G.723.1	6.3 Kbps	30 ms	5 ms	20 ms
MP-ACELP, G.723.1	5.3 Kbps	30 ms	5 ms	20 ms

5.2.2 Algorithmic Delay

The density algorithms trust to familiar voice characteristics to correctly process sample block N. The algorithm must need informations of what is in block N+1 in order to correctly repeat sample block N. It look ahead, which is actually an extra delay, is called algorithmic delay. This successfully increases the length of the compression block. This happens repeatedly, such that block N+1 looks into block N+2, and so forth and so on. The net effect is a 5 ms addition to the overall delay on the link. This means that the total time required to process a block of information is 10 m with a 5 ms constant overhead factor. For the examples in the remainder of this chapter, assume G.729 compression with a 30 ms/30 byte payload. In order to facilitate design, and take a conservative approach, the tables given in the remainder of this document assume the worst case coder delay. The coder delay, decompression delay, and algorithmic delay is lumped into one issue which is called the coder delay.(Cisco, 2000)

- Algorithmic Delay for G.726 coders is 0 ms
- Algorithmic Delay for G.729 coders is 5 ms.
- Algorithmic Delay for G.723.1 coders is 7.5 ms

The equation used to create the lumped Coder Delay Parameter is:

Equation 1 Lumped Coder Delay Parameter

(Worst Case Compression Time Per Block)
+
(De-Compression Time Per Block)
X (Number of Blocks in Frame)
+
(Algorithmic Delay)

= "Lumped" Coder Delay Parameter

Worst Case Compression Time Per Block: 10 ms

Decompression Time Per Block x 3 Blocks 3 ms

Algorithmic Delay 5 ms

Total (χ) 18 ms

5.2.3 Packetization Delay

Packetization delay is the required time to load in a total packet/cell before it is transmitted. Typically, G.711 pulse code modulation (PCM) encoded voice samples reach at the rate of 64 Kbps, which means it can take about 6 ms to fill the entire 48-byte payload of an ATM cell. The problem can be addressed either with partially filled cells or by multiplexing several voice calls into a single ATM virtual circuit channel (IEC, 2007).

Table 5.3 Common Packetization

Coder		**Payload Size (Bytes)**	**Packetization Delay (ms)**	**Payload Size (Bytes)**	**Packetization Delay (ms)**
PCM,	64	160	20	240	30

G.711	Kbps				
ADPCM, G.726	32 Kbps	80	20	120	30
CS-ACELP, G.729	8.0 Kbps	20	20	30	30
MP-MLQ, G.723.1	6.3 Kbps	24	24	60	48
MP-ACELP, G.723.1	5.3 Kbps	20	30	60	60

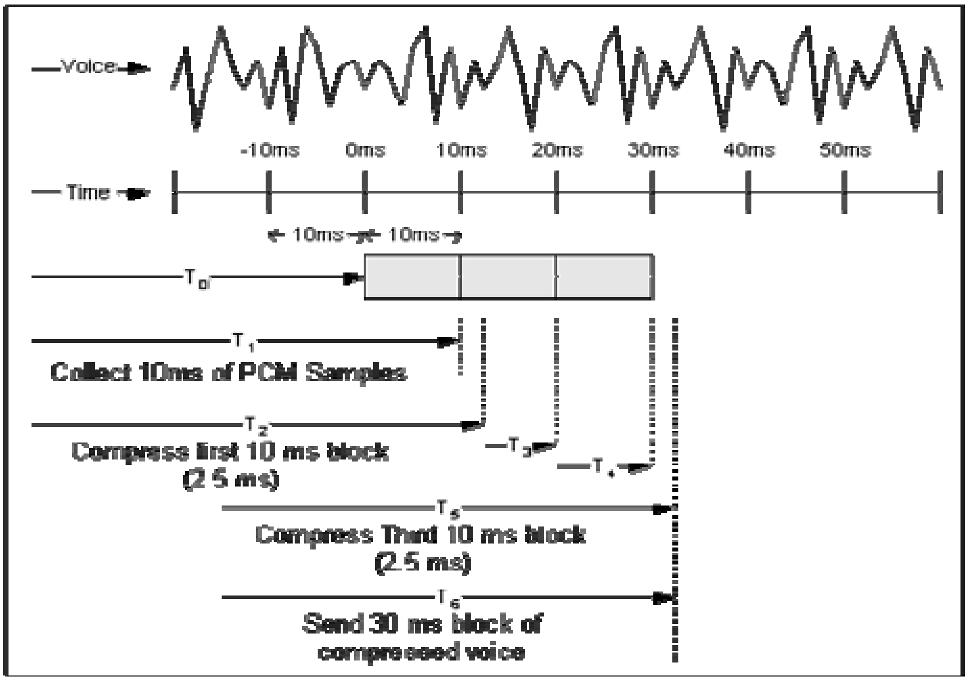


Figure 5.2 Pipelining and Packetization

The peak contour of the figure depicts a sample voice waveform. The second line is a time level in 10 ms increments. At T_0, the CS-ACELP algorithm begins to gather PCM patterns from the codec. At T_1, the algorithm has collected its first 10 ms block of samples and begins to squeeze it. At T_2, the first block of samples has been compressed. In this example the compression time is 2.5 ms, as indicated by T_2-T_1. The second and third blocks are collected at T_3 and T_4. The third block is compressed at T_5. The packet is assembled and sent (assumed to be instantaneous) at T_6. Due to the pipelined nature of the Compression and Packetization processes, the delay from when the process begins to when the voice frame is sent is T_6-T_0, or approximately 32.5 ms. For illustration, this example is based on best case delay. If the worst case delay is used, the figure is 40 ms, 10 ms for Coder delay and 30 ms for Packetization delay. (Cisco, 2000)

5.2.4 Serialization Delay

Serialization delay is the quantity of instant. It takes to essentially check a line and position the bits onto the chain for communication. This is an additional permanent structure of delay in networks. The delay will differ stand for the clocking tempo of the boundary. Clearly, a 56k switch contains an upper serialization delay than a T1 switch. The board below demonstrates the various delays with situation to the boundary speed.

Table 5.4 Serialization Delay

Frame Size (bytes)	Interface Speed (Kbps)							
	19.2	56	64	128	256	384	512	768
38	15.83	5.43	4.75	2.38	1.19	0.79	0.59	0.4
48	20	6.86	6	3	1.5	1	0.75	0.5
64	26.67	9.14	8	4	2	1.33	1	0.67
128	53.33	18.29	16	8	4	2.67	2	1.33
256	106.67	36.57	32	16	8	5.3	4	2.67
512	213.33	73.14	64	32	16	10.67	8	5.33
1024	426.67	149.29	128	64	32	21.33	16	10.67
1500	625	214.29	187.5	93.75	46.88	31.25	23.44	15.63
2048	853.33	292.57	256	128	64	42.67	32	21.33

5.2.5 Queuing/Buffering Delay

After the compressed voice consignment is built, a header is added and the frame is queued for communication on the network connection. Voice requires to have complete precedence in the router/gateway. For that reason, a voice frame must only wait for either a data frame that already plays out, or for other voice frames ahead of it. Fundamentally the voice frame waits for the serialization delay of any preceding frames in the output line. Queuing delay ($ß_n$) is a changeable delay and is reliant on the case speed and the state of the line. There are casual elements associated with the queuing delay. (Cisco, 2000)

5.2.6 Network Switching Delay

The public frame relay or ATM network that communicates the endpoint locations is the supply of the largest delays for voice lines. Network Switching Delays (ω_n) are the hardest to count as well. If wide area connectivity is provided, or some other private network, it is possible to identify the individual components of delay. In general, the fixed components are from propagation delays on the trunks within the network, and variable delays are from queuing delays clocking frames into and out of intermediate switches. In order to estimate propagation delay, a popular estimate of 10 microseconds/mile or 6 microseconds/km (G.114) is widely used. However, intermediate multiplexing equipment, backhauling, microwave links, and other factors found in carrier networks create many exceptions. The other major factor of delay is from queuing within the wide-area network. In a classified network, it can be probable to scale existing queuing delays or to guess a per hop account within the wide area network.(Cisco, 2000)

5.2.7 De-Jitter Delay

Speech is a steady bit rate ceremony, the delay variation from all the variable delays must be removed before the signal leaves the network. It is achieved with a de-jitter (Δ_n) buffer at the receiving end router or gateway. The de-jitter buffer converts the variable delay into a fixed delay. It holds the first sample received for a period of time before it plays it out. This time period is known as the primary play out delay. (Cisco, 2000)

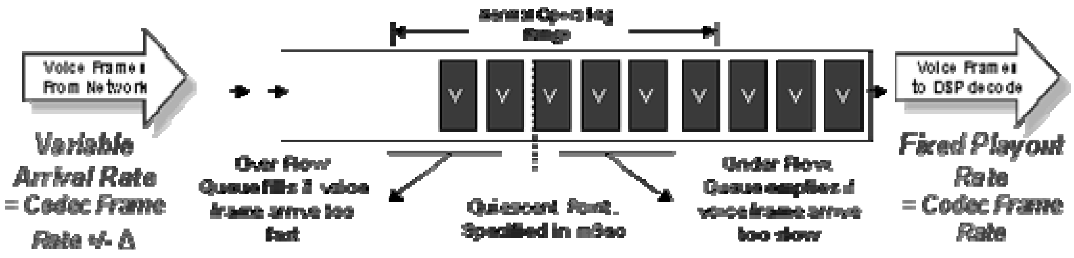


Figure 5.3 De-Jitter Buffer Operation

It is necessary to switch appropriately the de-jitter buffer . If samples are held for too short a time, variations in delay can potentially cause the buffer to under-run and cause gaps in the speech. If the sample is held for too long a time, the buffer can overrun, and the dropped packets again cause gaps in the speech. Lastly, if packets are held for too long a time, the overall delay on the connection can rise to unacceptable levels. The optimum initial play out delay for the de-jitter buffer is equal to the total variable delay along the connection. This is shown in Figure 5.4. The de-jitter buffers can be adaptive, but the maximum delay is fixed. When adaptive buffers are configured, the delay becomes a changeable form. However, the limit delay can be used as a worst case for devise intends. (Cisco, 2000)

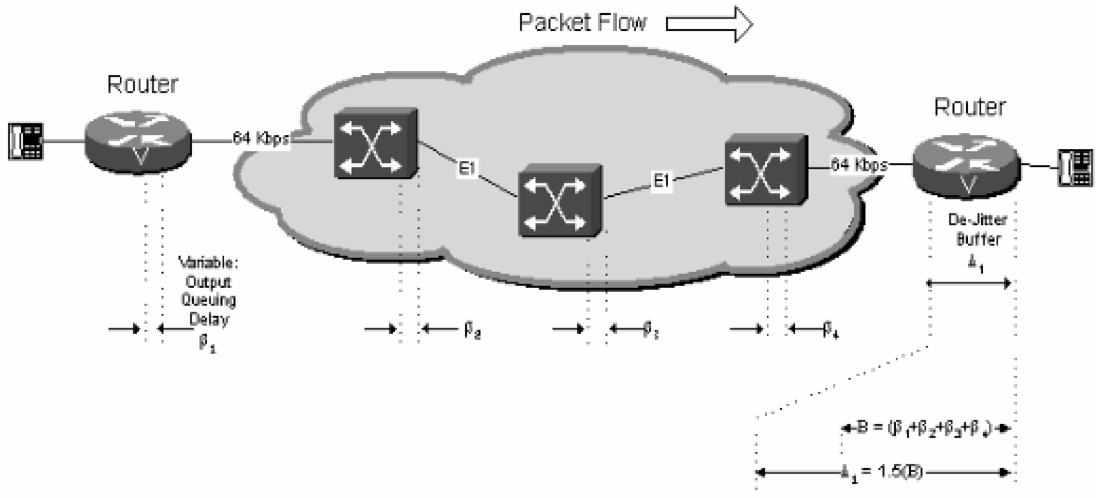


Figure 5.4 Variable Delay and the De-Jitter Buffer

The primary playout delay is editable. The highest depth of the buffer before it overflows is normally set to 1.5 or 2.0 times this value. If the 40 ms nominal delay setting is used, the first voice sample received when the de-jitter buffer is blank is held for 40 ms before it is played out. This implies that a following packet received from the network can be as much as 40 ms delayed (with respect to the first packet) without any loss of voice continuity. If it is delayed more than 40 ms, the de-jitter buffer empties and the next packet received is held for 40 ms before play out to reset the buffer. This results in a gap in the voice played out for about 40 ms. The actual contribution of de-jitter buffer to delay is the initial play out delay of the de-jitter buffer plus the actual amount the first packet was buffered in the network. The worst case is twice the de-jitter buffer initial delay (assumption is that the first packet through the network experienced only minimum buffering delay). In practice, over a number of network switch hops, it is probably not essential to suppose the worst case. (Cisco, 2000)

5.3 When Does Delay Become Too Much?

Delay impacts the atmosphere of a conversation instead of voice quality. under 100ms, lots of users will not realise the delay. Between 100ms and 300ms, users will realise a softly wavering in their partner's answer. This wavering can impact how each listener recognizes the mood of the talk. In this case, conversations can seem arctic. Disruptions are more recurrent, and the discussion gets out of hit. After the 300 ms, the delay is understandable to the users, and they begin to back off to avoid interruptions. At some point, conversation is almost not possible. Clearly, shorted delay effects better communication quality and better apparent overall voice quality.

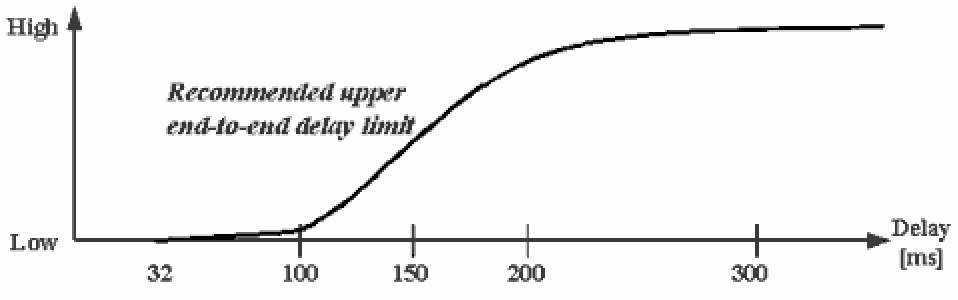


Figure 5.5 Delay's Effect On User Experience

5.4 Delay Effects On Voice Quality

G.114 standard suggests values for one way transmission time. ITU-T G.114 describes transmission period as the amount of propagation delay and processing delay. This describtion is alike of one way end to end delay or latency. The ITU-T recommends the followings for end to end delay:

Table 5.5 ITU-T Conditional Delay Acceptability

end to end delay (ms)	Acceptability
0 – 150	Acceptable for most applications
150 -400	Conditionally acceptable
Above 400	Generally unacceptable

The ITU G.114 standard stands that depending on the application. Delay not exceeding 150ms (one-way) is typically suitable. Delay between the 150 to 300 ms may or may not be satisfactory; if users hopes are lower. A higher delay rnay be acceptable. However, if one way delay goes beyond 400ms. The delay is too strict. The ITU-T recommends that the processing delay component of end to end delay not exceed 50ms. Several studies have been conducted on the effect of delay on the quality of voice communications. In a study reported in the ITU-T G.114 Recommendation the effect of pure deIay on the perceived quality of telephone connections was presented in the study. Delay of 0.250, and 500 ms were injected into the end-to-end delay of a telephone connection. The perceived quality of the connection was measured by a MOS survey. The results of the experiment are summarized in Table 5.6 below. (Sundstrom, 1999)

Table 5.6 MOS Rating of Delay

One way end to end delay (ms)	MOS Rating
0	good
250	fair
500	poor

As shown in table 5.6. delay breaks down the quality of the voice communication.

Table 5.7 Expected MOS Rating due to Delay

Continent	**Corresponding MOS Rating**
North America	Good
Asia	Fair
Australia	Fair
South America	Fair

Moreover to the challenges faced by conventional voice structure delay launchs further problems in packet switched voice systems. Large end to end delay may permit parties to attend to speaker echo. Speaker echo is reflected signal force caused by an impedance rnismatch in analogue telephony apparatus. Mechanism of a VoIP system may supply the chance to imitate signal force. For example, in a hybrid VoIP network, the voice call will get ahead of through various types of networks along the path from resource to target. If analogue telephony apparatus is present along the path, there may be a chance to turn out speaker echo. It has been shown that if end to end delay goes beyond 45-50 ms, speaker echo may be heard.(Sundstrom, 1999)

CHAPTER SIX
ECHO

Echo is a problem which is being progressively discerned in Voice over IP networks. In general echo is caused by mismatched hybrid on the analog part of a telephony connection. The other echo type is acoustic echo. It occurs because of feedback from speaker to microphone of a telephone handset.

Hybrid echo and acoustic echo are the two types of echo on voice networks. Hybrid echo is a linear electrical signal indication which occurs at the 4-wire to 2-wire conversion point in a PSTN network (usually found in a Class 5 PSTN switch). Hybrid echo can happen in the VoIP network where there is a connection between VoIP and PSTN networks. Non-linear one is acoustic echo and poor acoustic isolation between the speaker and the microphone of a user's device causes it. (e.g., handset, headset, softphone, speakerphone). Acoustic echo can enter the voice over IP network from any source.

With the added delay of the IP network, both types of echo become more obvious and annoying to the caller. Indeed, the added VoIP induced delay can make what would formerly be considered minor echo annoying that is enough to cause users to leave the call.

Commonly, VoIP Gateways integrate a line echo canceller to eliminate the echo level from analog loops. If it is not functioning truely, possible due the echo canceller being disabled, to mis-configuration of the signal levels (loss plan), non-linearity in the speech path or an excessively high echo level then some residual echo may be present. To resolve echo problems it is necessary to identify both the source of the echo (i.e. a particular analog loop or line card) and check its balance or configuration and then to know why the echo canceller is not sufficiently compensating for the echo. (Voiptroubleshooter, 2006)

According to a point of view, echo is the voice sound returning to the talker's ear via the speaker of the telophone. In other words, echo happens when the voice signal of the talker seep out from the transmit path back into the receive path.

Table 6.1 Relationship between Echo levels, delay and voice quality

One-Way Delay Range (ms)	Effect on Voice Quality
0–25	This is the expected range for national calls. There are no difficulties during conversation.
25–150	This is the expected range for international calls using a terrestrial transport link and IP telephony, which includes only one instance of IP voice. This range is acceptable for most users, assuming the use of echo control devices.
150–400	This is the expected range for a satellite link. Delays in this range can interrupt the normal flow of a conversation. A high-performance echo canceler must be used and careful network planning is necessary.
Greater than 400	This is excessive delay and must be avoided by network planning.

6.1 When Echo Becomes Perceptible?

As it is described before, roundtrip latency established into the voice path by VoP networks such as VoIP may frequently cause existing echo originating from an analog tail circuit to become perceptible and even annoying. Echo that originates between an individual's telephone and PSTN central office is not perceptible because it returns to one's ear too quickly. Even echo from the far-end tail circuit usually returns quickly enough or is attenuate enough to not be heard. Vop network componets, however, introduce into the voice path a fundemantel and unavoidable end to end delay that often exceedsthe 32 ms thresold mentioned earlier. If echo is produced in the far end PSTN analog tail circuit, at least twice this delay will pass before the echo reachs the near-end talkers ear. Because of this, even softed echo can become perceptible. As near end echo will not be heard, one can frequently and correctly finish that any perceptible echo creates from the far-end tail circuit. Figure 6.2 illustrates this point. (IEC,2007)

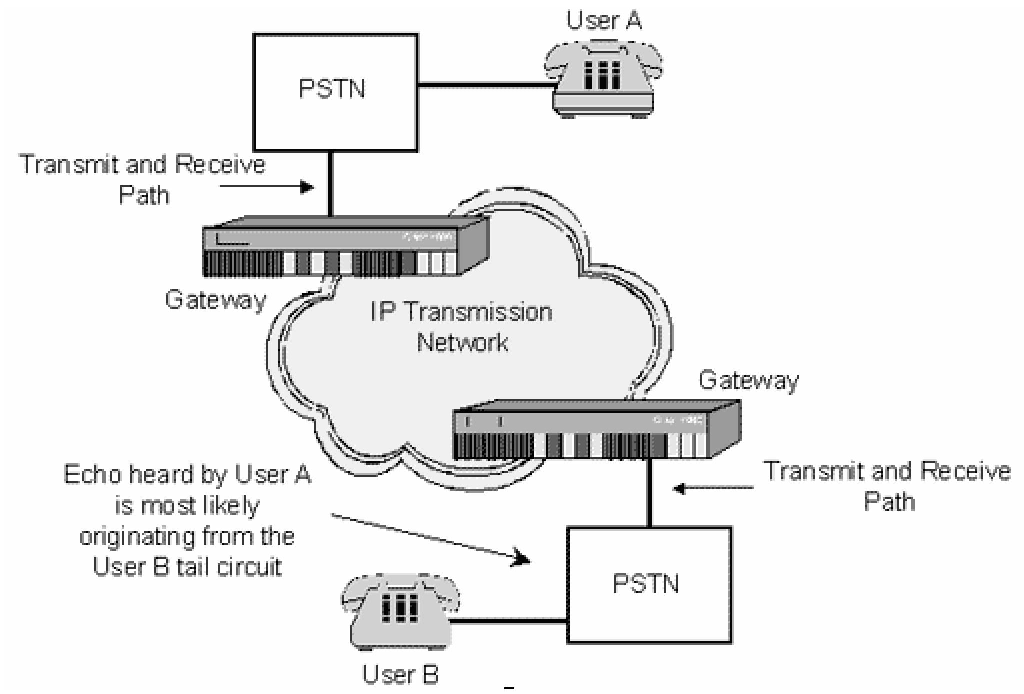


Figure 6.1 Echo originates from the far-end tail circuit

6.2 VoIP Networks And Echo

After Voice over IP echo troubles have been existed, the requirement of an IP based echo canceller has happen to apparent. As a result, most voice over IP networks require to echo cancellation. This is in dissimilarity to the PSTN where echo cancellation is only essential on long haul connections. In general, short delay echoes are rarely exceeds 30 ms. For this reason echo cancellation is not required on short PSTN connections. However, in voice over ip systems are unlikely to be less than 30 ms. It means that some form of echo cancellation is extremly required.

6.3 Acoustic Echo

While not as prevalent as echo caused by the hybrid (line echo), acoustical echo can also be encountered in the telecommunications networks. Acoustical echo is caused by poor isolation between the microphone and speaker of some telephone sets. Most hands free speakerphone systems incorporate special echo control circuitry to ensure that echo is not a problem. Another example is the need for acoustic echo cancellation to protect the landline subscriber from acoustic echo originating from digital wireless networks.

In the case of VoIP networks, acoustic echo is normally present when at least one of the callers is using a computer with a loudspeaker and a microphone. As is the case for line echo, acoustic echo becomes audible when there is long delay. On the other hand, differently from line echo, acoustic echo usually is not severe enough to make the conversation impossible. The methodology for canceling acoustic echo differs in many aspects from the methodology used for canceling line echo.

6.4 Line Echo

A difference on the impedance from the four-wire network switch conversion to the two-wire local loop causes line echo in PSTN. The 2-wire local loop consists of a single pair of wires that carry both directions of the conversation. At the local telephone exchange, this 2-wire pair is connected to a 4-wire trunk by using a device called a hybrid. The 2-wire local loop splitted by the hybrid into two distinct pairs of wires, one for the send path and one for the receive path as described by the following figure.

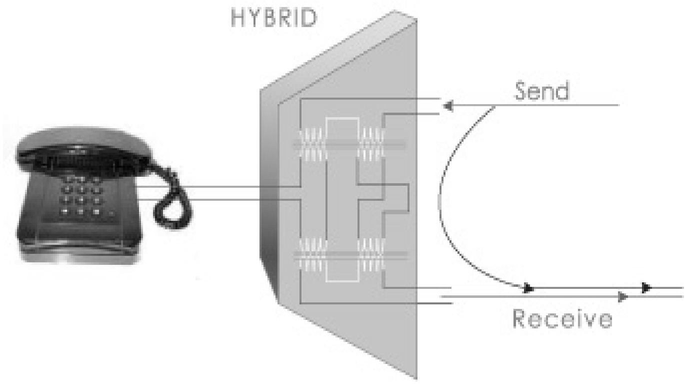


Figure 6.2 The hybrid device and line echo generation

Because the hybrid cannot be made to split the 2-wire loop perfectly, some of the receive signal is erroneously leaked into the send path and is called echo. This mechanism is not best to remove echo. For example, because of the echo suppressor discontinues the frequency range that is used by the ISDN, a line that has an echo suppressor can not use Integrated Services Digital Network. The proposed algorithm does not deal with echo suppressors.

On the other hand, in IP networks, echo cancellers can be built into the codecs and operate on each DSP. In proposed algorithm it takes advantage of the measurements made by those echo cancellers present in the DSP to draw conclusions about the echo quality in the call and more generally the voice quality of the call. It should be noted that once the echo canceller has already computed such measurements there is no extra computational effort required by the algorithm for the DSP.

At the following figure, there is a very simple illustration of a TDM-IP gateway with a line echo canceller.

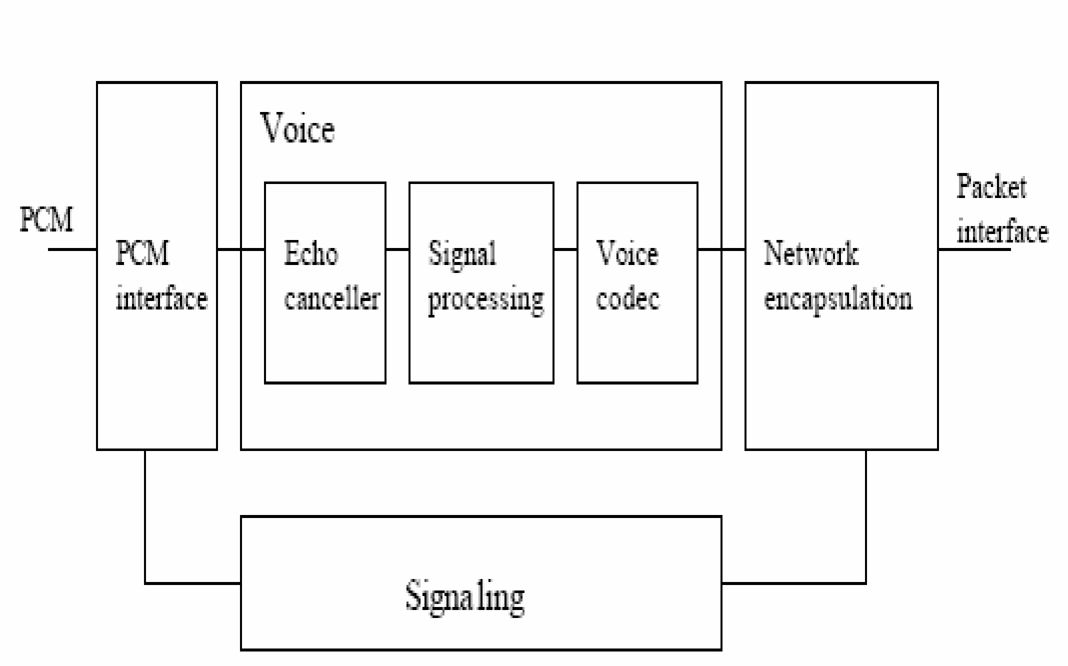


Figure 6.3 Simplified block diagram of a TDM-IP gateway

6.5 Echo Cancellation

Echo cancellation is the operation of removing echo from a voice communication to improve the quality of the voice call. Echo cancellation is frequently required since speech compression techniques and packet processing delays generate echo. There are 2 types of echo: acoustic echo and line echo.

The level to which echo is objectionable relies on total delay and echo loudness. The total delay is combined with the process of digitally encoding the voice, delay in digital processing on both ends (for instance time slot assignment or packetizing, in the case of data-oriented circuits), and two times any delay in the long distance circuitry.

At the Figure 6.5, ITU-T G.131 recommendation can be observed and it describes the variation in acceptable perceived echo loudness versus total delay,

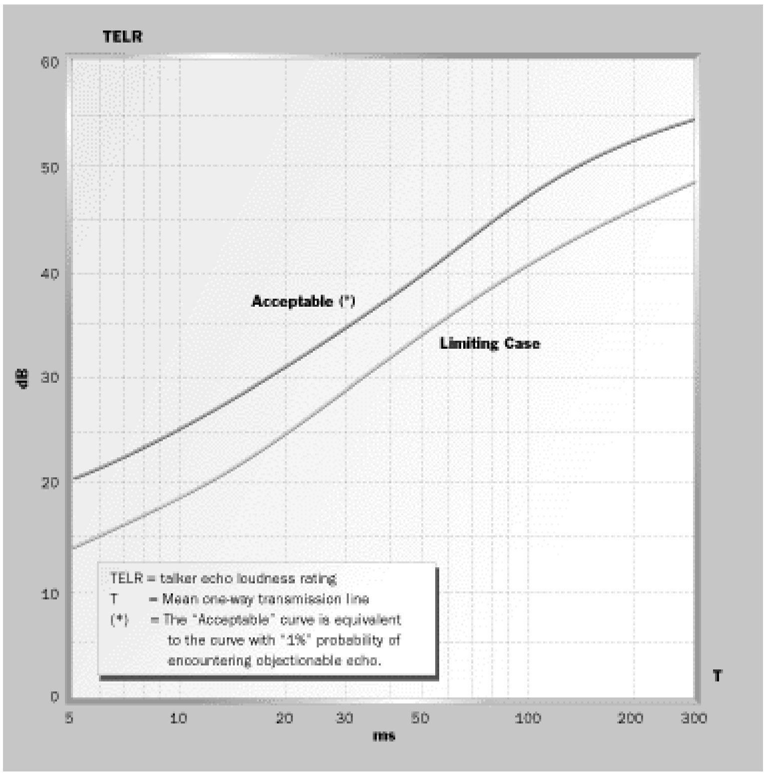


Figure 6.4 Talker echo tolerance curves

TELR = talker echo loudness rating = SLR + RLR + R + T + Lr

SLR = speaker loudness rating = 7dB nom, 2dB min for most telephones

RLR = receiver loudness rating = 3 dB nom, 1 dB min for most telephones

R = receive loss

T = transmit loss

R+T = 6 dB is introduced in most calls in the US for echo control

Lr = return loss or hybrid balance = 14 dB nom, 8 dB min for line cards that subtract a constant fraction of the signal being sent due to the variation in telephone and loop impedance. (Gamel, 2003)

The four ports of the echo canceller are denoted as follows:

• Receive- in (Rin)
• Receive-out (Rout)
• Send-in (Sin)
• Send-out (Sout).

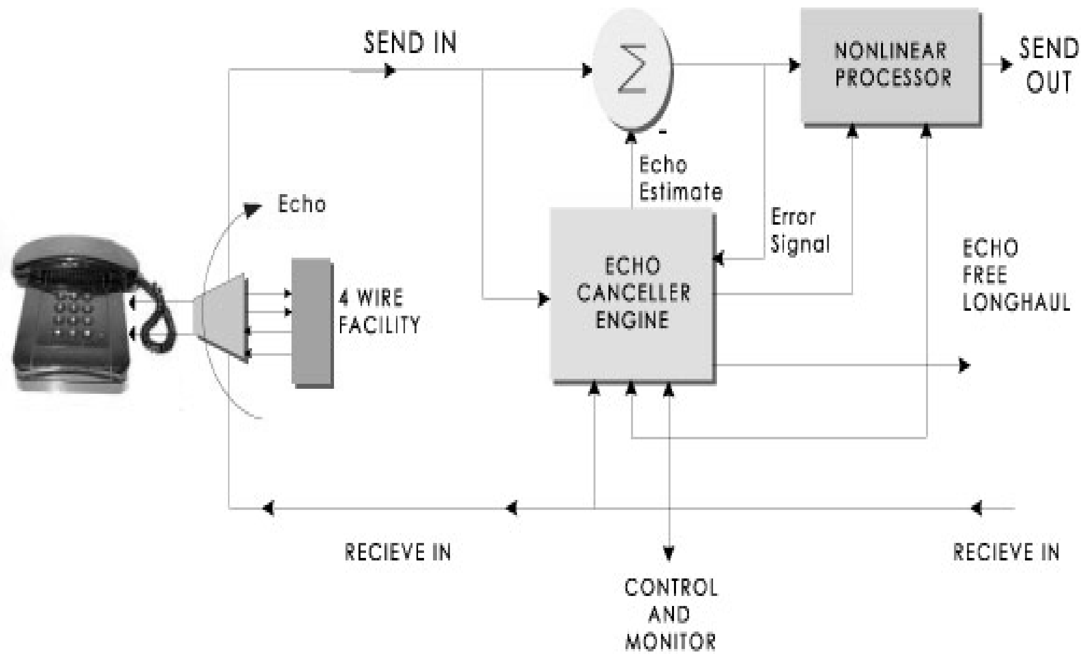


Figure 6.5 Circuit with a line echo canceller

Speech is observed by an echo canceller from the far end that passes through its receive path and uses this information to compute an estimate of the echo that is then subtracted from its send path. If the estimation is good, the echo is cancelled and only the near end speech is sent to the far end. Good echo cancellation is essential for the quality of the voice in the network. Echo cancellation occurs between the send-in and send-out ports, reducing the echo present in the send path. The total amount of echo attenuation that an echo canceller provides is called echo return loss enhancement (ERLE). ERLE is the difference in the echo level between the send- in and send-out ports and it is measured in dB.(Kauffman, 2006)

An echo canceller normally consists of three major building blocks:

- Adaptive filter
- Double-talk detector
- Nonlinear processor

The next figure is improved version of the echo canceller that was represented in Figure 6.6 with its major building blocks listed above.

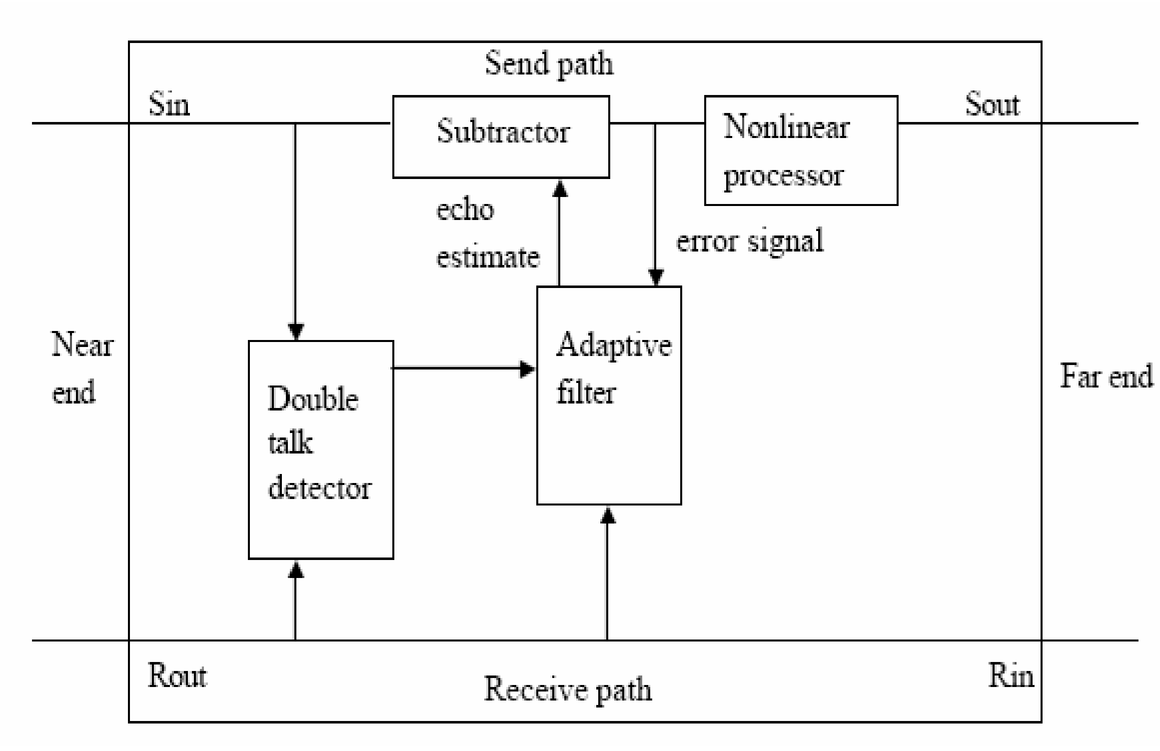


Figure 6.6 Block diagram of a line echo canceller

The quantity of echo attenuation supplied by the hybrid. That is, the attenuation of the signal from the Rout port to the Sin port of the echo canceller.

Echo cancellation occurs when the signal coming from the Rin port is sampled and given to the adaptive filter. The signal then travels from the Rout port of the echo canceller, to the hybrid, where most of the signal is transferred to the 2-wire loop connected to the near-end telephone. A portion of the signal is leaked by the hybrid to the Sin port of the echo canceller. This is the echo that needs to be cancelled by the adaptive filter. The echo path (Figure 6.4) is highly variable, so the filter that is required to realize the echo cancellation can not be a fixed filter. In fact, the echo path must be estimated for the particular local loop to which the hybrid gets connected. One option to derive the filter is to measure the impulse response of the echo path and then approximate it by a tapped delay line. However, in general the echo path is not stationary. Therefore, such measurements would have to be made repeatedly during a conversation. To eliminate the need of such measurements the filter is made adaptive. An algorithm is implemented which uses the residual error to adapt the filter to the characteristics of the local loop. The adaptive filter computes an estimate of the echo. The resulting estimation of the echo is then subtracted from the signal coming from the Sin port, which is composed by the echo and possibly some near end speech and noise.(Kauffman, 2006)

The resulting output is residual echo that is passed on to the nonlinear processor and is also fed back to the adaptive filter as the error signal. However, this error signal is truly an error signal only when there is no near end speech. If there is near end speech, the “error signal” does not accurately indicates the degree of success of the cancellation and the adaptation algorithm will not converge, resulting in a failed attempt to cancel the echo. For this reason, there is a need to have double talk detection, so that the adaptation would only occur when there is no double talk. When the echo canceller’s double talk detector senses that both the near end and far end callers are speaking at the same time, it informs the adaptive filter so that the filter can ignore the error signal that comes from the subtractor, freezing the filter adaptation. As we said before, near end speech can distort the error signal and confuse the adaptation process, for this reason adaptation is halted when double talk is detected. Of course, the echo canceller still continues to cancel echo during doubletalk. As soon as the double talk detector senses that double talk is no longer present, it informs the adaptive filter so that it can, once again, use the error signal to adapt to the impulse response of the hybrid.

The quantization noise introduced by the PCM representation of speech samples and nonlinear echoes make it difficult for the adaptive filter to develop an absolutely perfect echo estimate. Nonlinear echoes can be caused by clipped speech signals, speech compression or poor quality speakerphones. It is extremely difficult to develop an accurate echo estimate of these nonlinear echoes because the echo canceller’s linear impulse response model cannot be correlated with these nonlinear echoes. Consequently, residual echo from the subtractor is reduced to an inaudible level by some nonlinear processing. The nonlinear processor has a suppression threshold that is typically adaptive, based on the Rin and Sin signal levels. The threshold is made adaptive because, if the nonlinear processor simply blocked all signals in the send path, there would be noticeable clipping of speech. For a more detailed description about a nonlinear processor.(Kauffman, 2006)

CHAPTER SEVEN
JITTER

jitter is the difference of packet interarrival time. It is one subject that exists only in packet based Networks and it is a variation in packet transit delay caused by queuing, contention and serialization effects on the path through the network. In general, higher levels of jitter are more likely to occur on either slow or heavily congested links. It is expected that the increasing use of “QoS” control mechanisms such as class based queuing, bandwidth reservation and of higher speed links such as 100 Mbit Ethernet, E3/T3 and SDH will reduce the incidence of jitter related problems at some stage in the future, however jitter will remain a problem for some time to come.(Voiptroubleshooter, 2000)

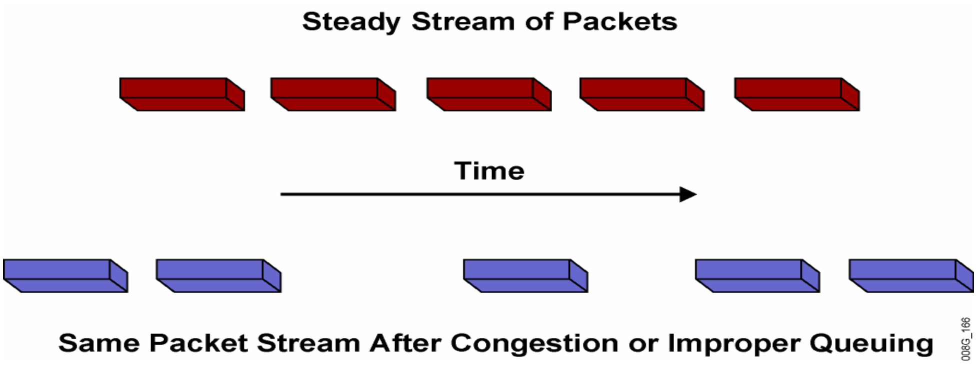


Figure 7.1 Packet Stream with Congestion

Jitter is described as the variability of delay suffered by different packets. The voice sounds journey by a few different route, the total quantity of delay experienced by following packets can differ to the point where the order in which data is transrnitted differs from the order in which it is received. For standard data, it does not characterize much of a problem. However, when dealing with speech, the data must be processed in a continuous manner such that the spacing between successive samples after decoding is identical to that of the transmitted signal. Jitter is then unacceptable for speech and must be properly dealt with to ensure good speech quality in a voice over internet protocol system. The part of the voip system that deals with the jitter is the playout buffer.

Easy method to understand the role of the playout buffer is to consider an example where the delay introduced by the playout buffer is fixed. Consider a playout buffer of delay 100 ms and packets that contain 10 ms of speech. To achieve good quality speech, a new speech packet must be played every 10 ms otherwise it will have a gap of silence. Suppose further that received packet 1 at 10 ms, packet 3 at 80 ms and packet *2* at 100 ms. Note that the number associated with the packet indicates the order they were transmitted and the order in which they should be played at the destination to reproduce the correct speech. If no buffer is used, it would only start playing packet 1 when it arrives at 10 ms, causing an initial silence of 10 ms. We would then play packet 3 at 80 ms. causing a gap of silence of *60* ms, followed by packet *2* at 100 ms, causing an additional gap of silence of 10 ms. Taking into account the amount of silence introduced that was not initially there and the fact that the speech was not played in the right order, the quality of the speech for a system with no buffer is clearly unacceptable. Now reconsider the same case with a playout buffer with a delay of 100 ms. Packet 1 is received at 10 ms, but it will not be played before 100 ms due to the playout buffer. At 80 ms, it receive packet 3 and simply buffer it since we do not need to play it at that time. At 100 ms, it receive packet *2* and the speech of packet 1 is played. At 110 ms, it play packet 2 and at 120 ms, it play packet 3 which was stored in the buffer. The result is speech that has been delayed by 100 ms but is perfectly intact compared to the original. This instance obviously specify the intent of a playout buffer and why it is necessary in a VoIP procedure. (Montminy, 2000)

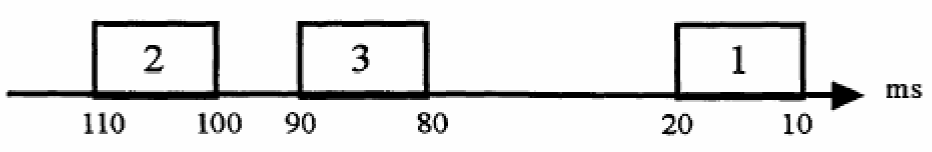


Figure 7.2 Packet Arrival Example

It is clear from this instance, the delay launched by the playout buffer is necessaryly the maximum amount of delay that is allowed for a packet in the system. If a packet arrives past this maximum amount of delay, it wil not be considered and the system will need to "play" something in its place in order to maintain the constant playout interval. The playout buffer essentialy removes all the jitter of the system at a cost of additional delay and dropping packets that arrive too late. The amount of delay added by the playout buffer is chosen large enough to ensure that most packets have time to arrive but small enough to avoid adding too much delay to the system. To optimize the performance of the playout buffer, the size of the buffer can be made adaptive. By tracking the delay suffered by each packet through a scheme such as RTCP, the buffer can adjust its delay to minimize the overall delay introduced and maximize the number of packets received on time. In order that the change of playout buffer delay not be noticeable, it is usually done progressively and only during periods of silence in the speech. Adaptive playout buffers use algorithm such as LMS to optimize performance. Since jitter is unacceptable for speech, the use of an adaptive playout buffer is essential in a VoIP system and translates the problem of jitter into a problem of delay and packet loss. (Montminy, 2000)

7.1 Jitter Measurement

Different techniques have been used to measure jitter rates but there are no good representations to supply beter results of the three types of jitter which described above.

7.1.1 Packet to Packet Jitter

Real time control protocol jitter measurement uses packet to packet jitter as a foundation. If the delay of this two consecutive packets is t1 and t2, Its called the packet to packet jitter is abs (t2-t1). The quantity estimated using this method matchs to the peak to peak jitter level only if the packets arrive alternately early and late. For example, if packets arrived according to the following sequence early, early, late, late then the reported value would be half that for the sequence early, late, early, late. It should be noted that in the case of RTCP, the value reported only reflects the most recent few hundred milliseconds before the value was calculated. If an RTCP report is sent every 10 seconds then no helpful knowledge is accessible interesting over 95% of the time between reports. (Voiptroubleshooter, 1)

7.1.2 Absolute packet Jitter

If the supposed coming period (denoted below ai) for a packet is known or can be determined then the absolute delay variation is abs(ti – ai). That rate can be misleading if a delay alter happens, as a stable offset would be involved. As even fixed delay variation buffers can settle to delay shifts this means that the reported jitter value would not necessarily be a good indicator of ideal jitter buffer size or discard rate. Another approach is to control the mean absolute packet delay variation with regard to a short term average or minimum value termed here the adjusted absolute packet delay variation. It can supply significant connection to delay variation buffer behavior. (Voiptroubleshooter, 1)

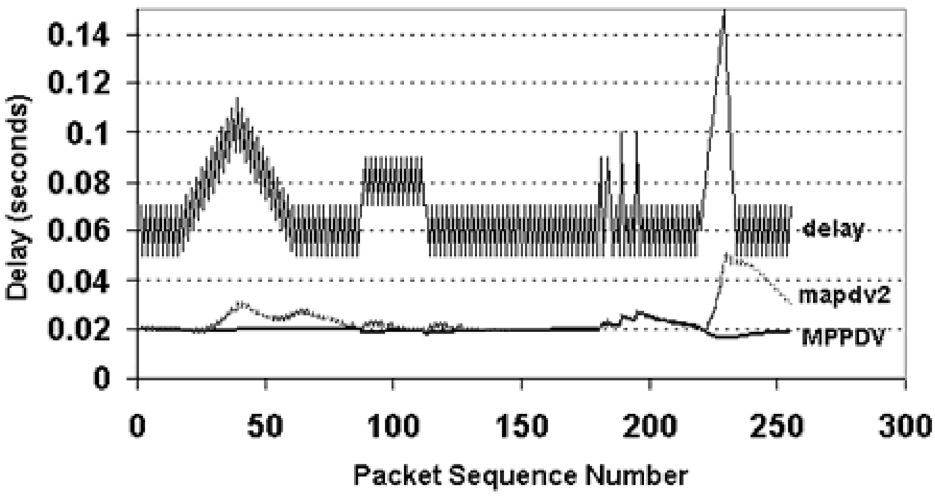


Figure 7.3 Comparison of Running Average Packet-to-Packet and Adjusted Absolute Delay Variation values for simulated jitter distribution

At the absolute packet Jitter, MPPDV and MAPDV2 answer in the same way to steady levels of jitter and to high variability in delay but MPPDV does not become aware of the ramp like delays feature of entrance link blockage.

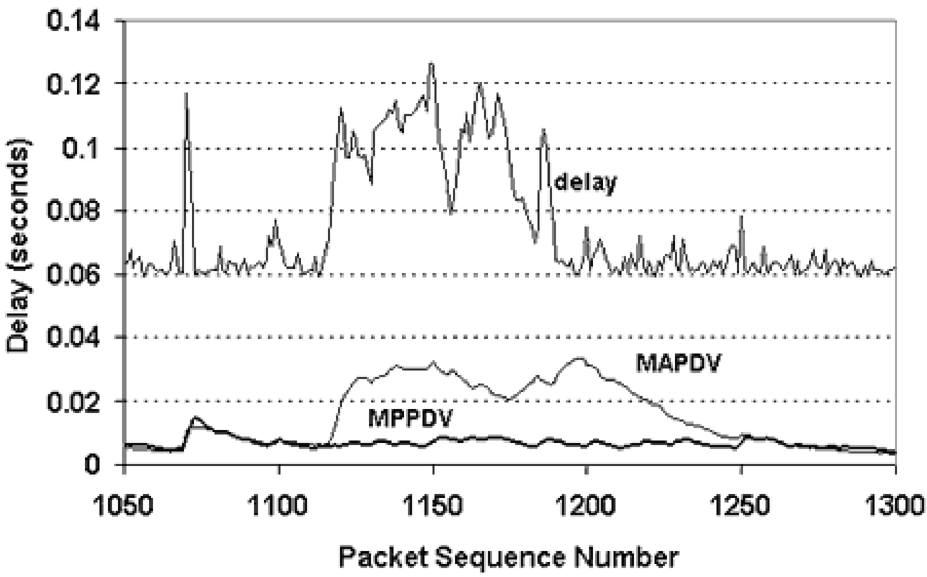


Figure 7.4 Comparison of Running Average Packet-to-Packet and Adjusted Absolute

7.2 Jitter Prevention

7.2.1 Anti-Jitter Circuits(AJCs)

AJCs are a group of chips designed to decrease the intensity of jitter in a usual pulse signal. AJCs control by retiming the output pulses so they align more closely to an idealised pulse signal. They are widely used in clock and data recovery circuits in digital communications, as well as for data sampling systems such as the analog-to-digital converter and digital-to-analog converter. Examples of anti-jitter circuits include phase-locked loop and delay-locked loop. Inside digital to analog converters jitter causes unwanted high-frequency distortions. In this case it can be suppressed with high fidelity clock signal usage.(Wikipedia, 2006)

7.2.2 Jitter Buffer

A jitter buffer is a collective information region where voice packets can be composed, stocked up, and launched to the voice CPU in calmly spaced gaps. It is refered a jitter filter, a mechanism or software method that abolishs jitter caused by communication delays in an Internet telephony (VoIP) network. Differences in packet coming time, refered jitter, may happen due to network blocking, timing float, or path transforms. The jitter buffer, which is positioned at the destination of the voice link, purposely delays the incoming packets so that the destination user understands an obvious correlation with too small voice deformation.

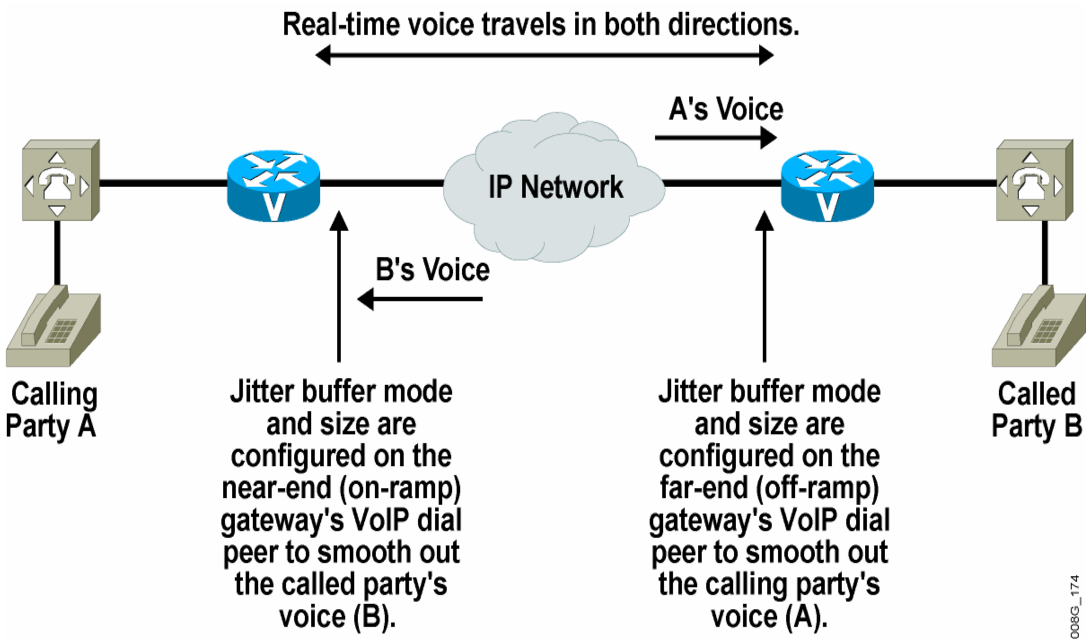


Figure 7.5 Jitter Buffer Operation

There are two sorts of jitter buffers. The first one of them is static jitter buffer and other one is dynamic jitter buffer. A static jitter buffer is hardware based and is organized by the producer. A dynamic jitter buffer is software based and may be organized by the network managers to become accustomed to alters in the network's delay.

Jitter buffers or dejitter buffers are used to defy jitter set up by packet networks so that a permanent playout of audio (or video) transmitted over the network can be ensured. The highest jitter that can be countered by a de-jitter buffer is equivalent to the buffering delay introduced earlier than initial the play-out of the mediastream. (Wikipedia, 2006)

CHAPTER EIGHT
VOICE QUALITY TESTS

Mean opinion score is the most well known measure of voice quality. Because of this, tests were done using MOS rates. Voice over ip network, factors that affect the convergence, voice quality and solutions to ensure a good conversation quality were explained before. These tests were done to compare packet loss rates and how to impact the conversation, jitter buffer and it's influence. 16 tests were applyed include the different codecs with different bit rates. Besides, some tests iclude voice activity detections and packet loss concealments to rate the conversation quality.

8.1 VOIP Quality Tests

Packet loss concealment and voice activity detection ensure a better conversation quality. At first test, G723.1 codec was used and observed how to impact network with various jitter rates, packet loss, voice activity detection and packet loss control (PLC) .

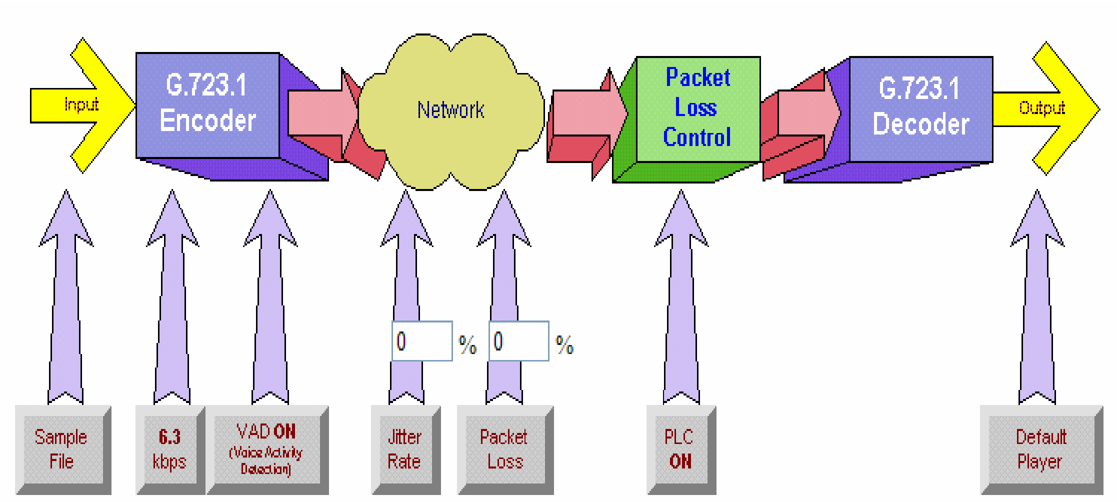


Figure 8.1 Test graph

1) For G723.1, 6.3 kbps, VAD is on, Jitter rate is %0, Packet loss rate is %0 and PLC is on,

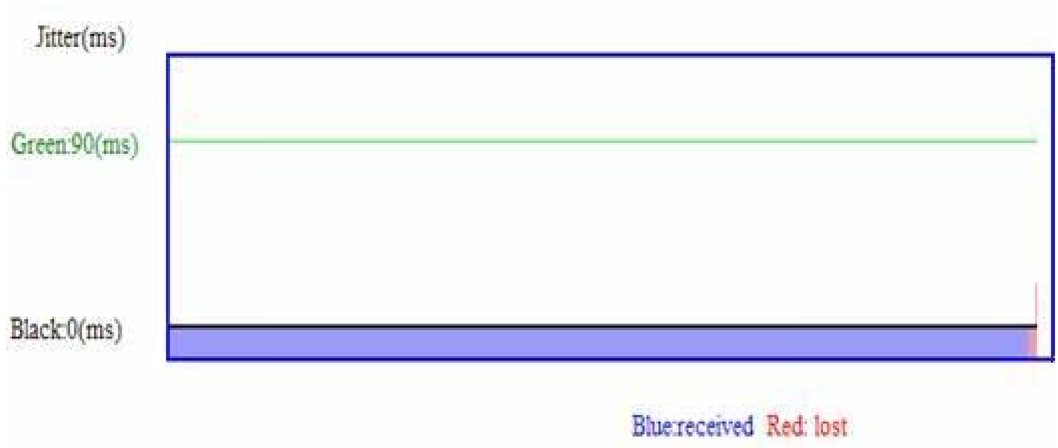


Figure 8.2 Test graph result

Table 8.1 Test results

Network Statistics					
Total Packets: 6.3 kbps Packets: 5.3 kbps Packets: Silence Packets: Empty Packets: Unknown Packets:	525 516 0 4 5 0	Packets lost: Jitter exceed 90 ms: Maximum: Mean: Variance:	0 0 0 0 0.000	Packets late (discarded): Packets not found: Total packet lost: Total Packet loss %:	0 0 0 0.00

It is clear on the figure that there is no packet loss and latency on the first test and listener's opinion will be good.

2) For G723.1, 6.3 kbps, VAD is on, Jitter rate is %5, Packet loss rate is %0 and PLC is on,

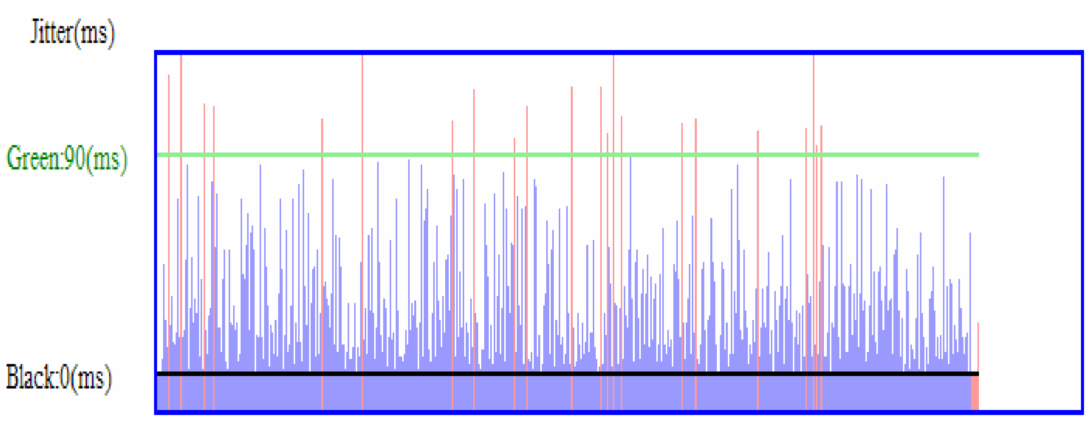


Figure 8.3 Test graph result

Table 8.2 Test results

Network Statistics					
Total Packets: 6.3 kbps Packets: 5.3 kbps Packets: Silence Packets: Empty Packets: Unknown Packets:	525 516 0 4 5 0	Packets lost: Jitter exceed 90 ms: Maximum: Mean: Variance:	0 22 237 30 1.280	Packets late (discarded): Packets not found: Total packet lost: Total Packet loss %:	22 22 22 4.19

At the second test, because of the jitter effects, some voice samples were lost and it caused to a decrease on the conversation quality.

3) For G723.1, 6.3 kbps, VAD is on, Jitter rate is %0, Packet loss rate is %5 and PLC is on,

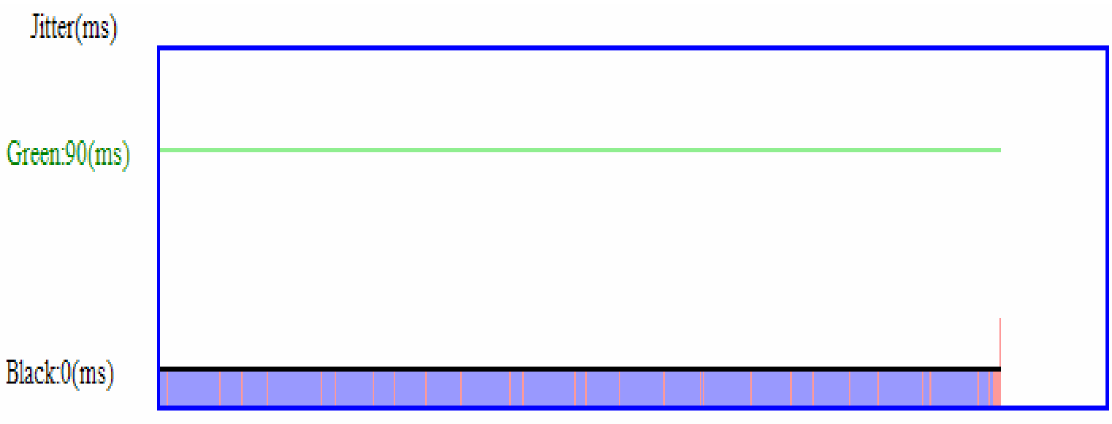


Figure 8.4 Test graph result

Table 8.3 Test results

Network Statistics						
Total Packets: 6.3 kbps Packets: 5.3 kbps Packets: Silence Packets: Empty Packets: Unknown Packets:	525 516 0 4 5 0	Packets lost: Jitter exceed 90 ms: Maximum: Mean: Variance:	27 0 0 0 0.000	Packets late (discarded): Packets not found: Total packet lost: Total Packet loss %:	0 27 27 5.14	

There is approximately %5 packet loss rate on this test but because of the packet loss concealment, it couldnt be noticed by listeners.

4) For G723.1, 6.3 kbps, VAD is on, Jitter rate is %5, Packet loss rate is %5 and PLC is on,

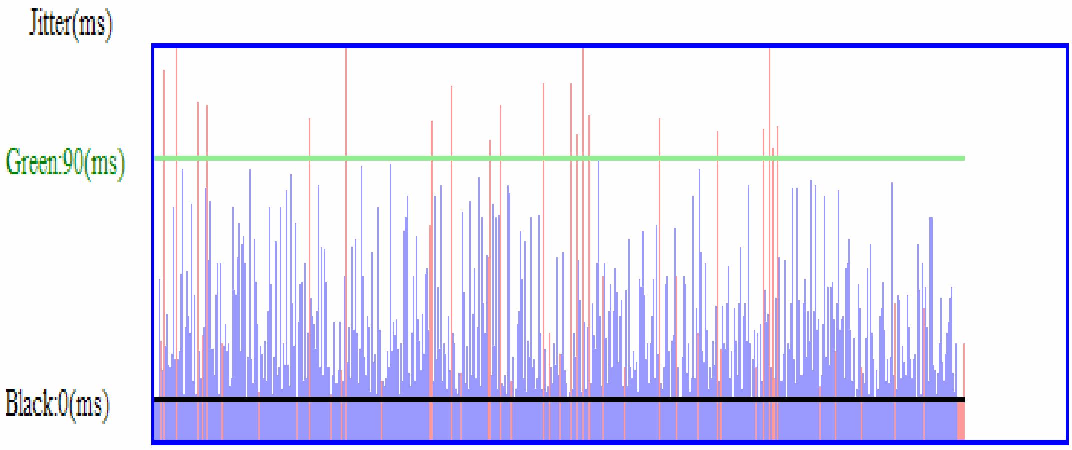


Figure 8.5 Test graph result

Table 8.4 Test results

Network Statistics						
Total Packets: 6.3 kbps Packets: 5.3 kbps Packets: Silence Packets: Empty Packets: Unknown Packets:	525 516 0 4 5 0	Packets lost: Jitter exceed 90 ms: Maximum: Mean: Variance:	26 21 237 30 1.268	Packets late (discarded): Packets not found: Total packet lost: Total Packet loss %:	21 47 47 8.95	

There are both jitter and packet loss impairments at this test. Conversation quality decreased at a noticable rate.

5) For G723.1, 6.3 kbps, VAD is off, Jitter rate is %5, Packet loss rate is %5 and PLC is on,

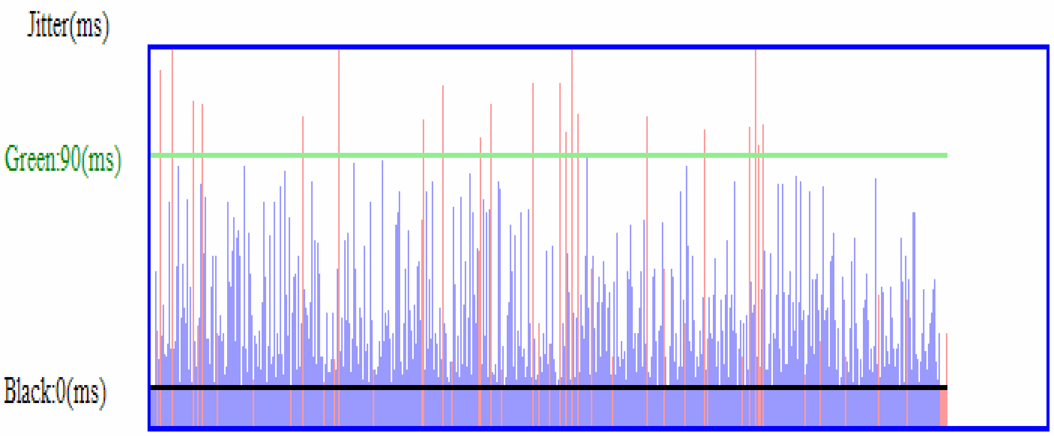


Figure 8.6 Test graph result

Table 8.5 Test results

Network Statistics						
Total Packets: 6.3 kbps Packets: 5.3 kbps Packets: Silence Packets: Empty Packets: Unknown Packets:	525 525 0 0 0 0	Packets lost: Jitter exceed 90 ms: Maximum: Mean: Variance:	26 21 237 30 1.268	Packets late (discarded): Packets not found: Total packet lost: Total Packet loss %:	21 47 47 8.95	

Voice activity detection was off mode and the conversation quality decresed because of VAD, jitter and packet loss.

6) For G723.1, 6.3 kbps, VAD is on, Jitter rate is %5, Packet loss rate is %5 and PLC is off,

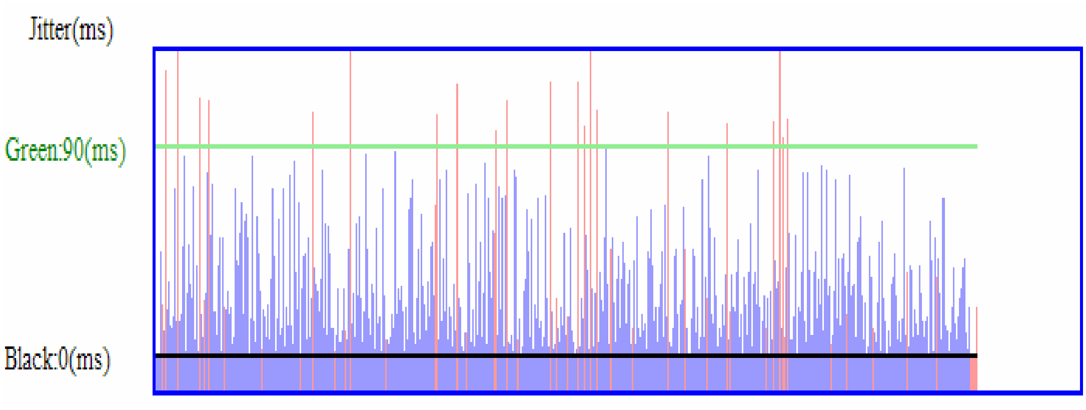


Figure 8.7 Test graph result

Table 8.6 Test results

Network Statistics						
Total Packets: 6.3 kbps Packets: 5.3 kbps Packets: Silence Packets: Empty Packets: Unknown Packets:	525 516 0 4 5 0	Packets lost: Jitter exceed 90 ms: Maximum: Mean: Variance:	26 21 237 30 1.268	Packets late (discarded): Packets not found: Total packet lost: Total Packet loss %:	21 47 47 8.95	

For the other tests, G726 codec was used and observed muLaw and ALaw between 40 kbps and 16 kbps.

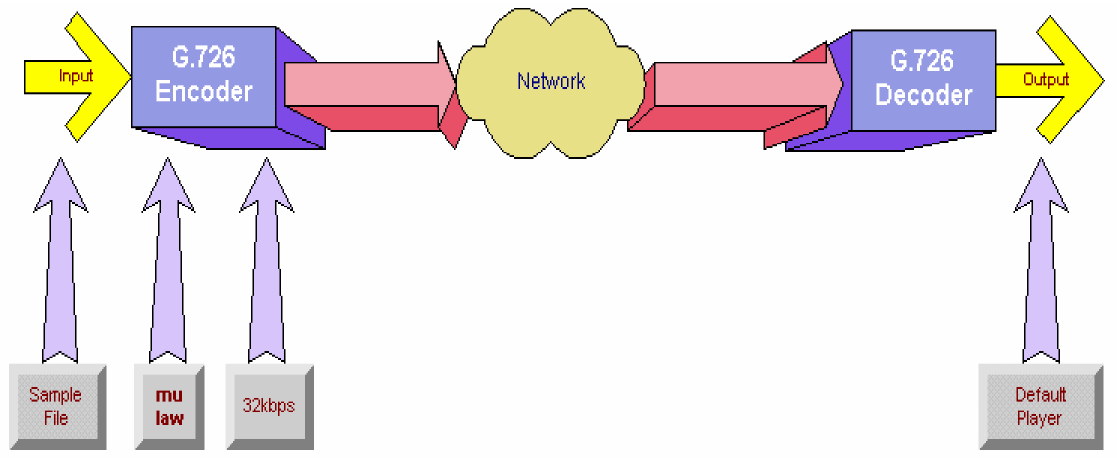


Figure 8.8 Test graph

7) G726, 40 kbps, muLaw ,

8) G726, 16 kbps, muLaw,

9) G726, 40 kbps, ALaw,

10) G726, 16 kbps, ALaw

At last test, G729 codec was used and observed jitter, packet loss, voice activity detection and packet loss control (PLC) again.

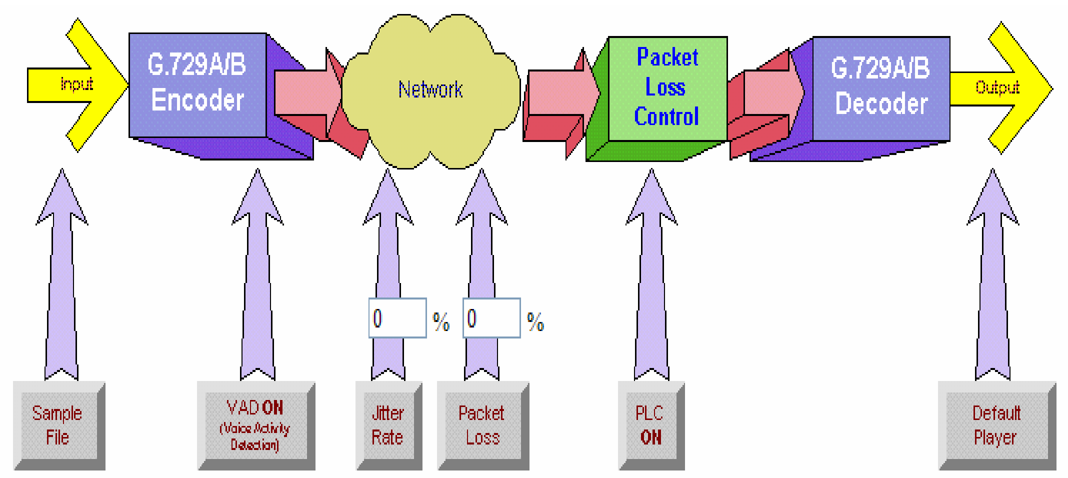


Figure 8.9 Test graph

11) For G729, 8 kbps, VAD is on, Jitter rate is %0, Packet loss rate is %0 and PLC is on,

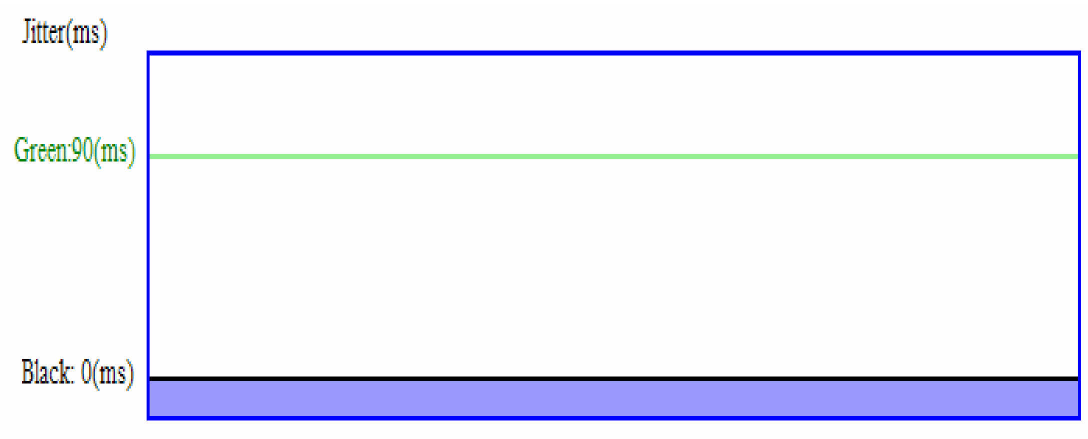


Figure 8.10 Test graph result

Table 8.7 Test results

Network Statistics						
Total Packets: 8 kbps Packets: Silence Packets: Empty Packets: Unknown Packets:	1575 1455 44 76 0	Packets lost: Jitter exceed 90 ms: Maximum: Mean: Variance:	0 0 0 0 0.000	Packets late(discarded): Packets not found: Total packet lost: Total Packet loss %:	0 0 0 0.00	

12) For G729, 8 kbps, VAD is on, Jitter rate is %5, Packet loss rate is %0 and PLC is on,

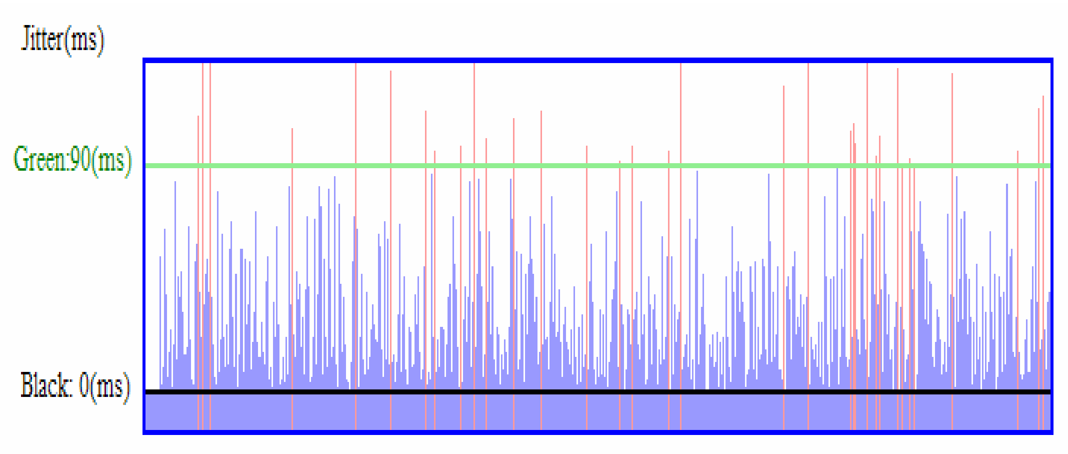


Figure 8.11 Test graph result

Table 8.8 Test results

Network Statistics						
Total Packets: 8 kbps Packets: Silence Packets: Empty Packets: Unknown Packets:	1575 1455 44 76 0	Packets lost: Jitter exceed 90 ms: Maximum: Mean: Variance:	0 83 204 30 1.154	Packets late(discarded): Packets not found: Total packet lost: Total Packet loss %:	83 83 83 5.27	

13) For G729, 8 kbps, VAD is on, Jitter rate is %0, Packet loss rate is %5 and PLC is on,

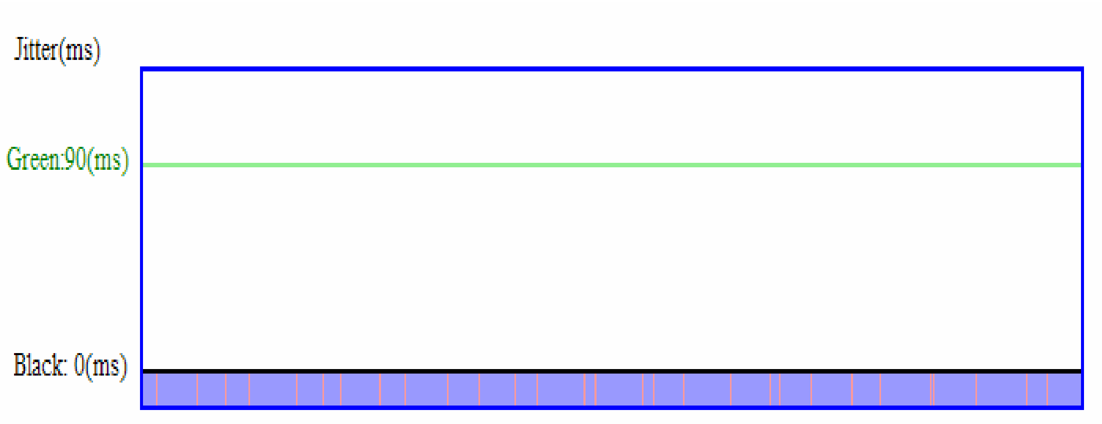


Figure 8.12 Test graph result

Table 8.9 Test results

Network Statistics						
Total Packets: 8 kbps Packets: Silence Packets: Empty Packets: Unknown Packets:	1575 1455 44 76 0	Packets lost: Jitter exceed 90 ms: Maximum: Mean: Variance:	79 0 0 0 0.000	Packets late(discarded): Packets not found: Total packet lost: Total Packet loss %:	0 79 79 5.02	

14) For G729, 8 kbps, VAD is on, Jitter rate is %5, Packet loss rate is %5 and PLC is on,

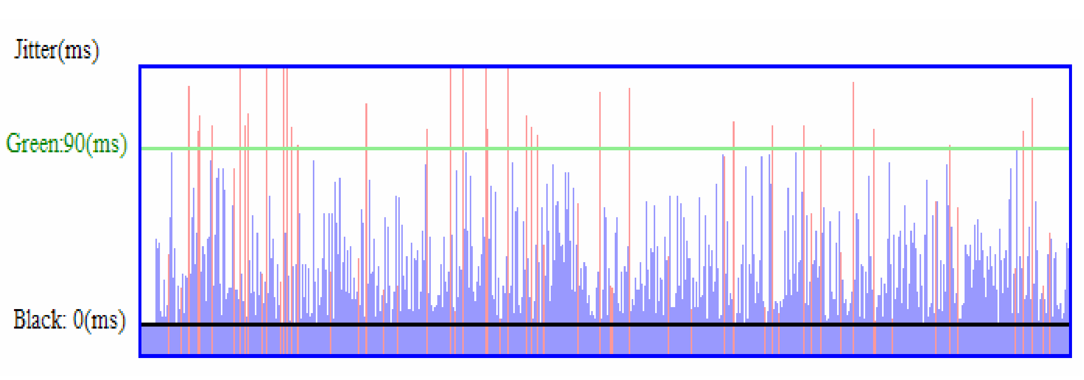


Figure 8.13 Test graph result

Table 8.10 Test results

Network Statistics						
Total Packets: 8 kbps Packets: Silence Packets: Empty Packets: Unknown Packets:	1575 1455 44 76 0	Packets lost: Jitter exceed 90 ms: Maximum: Mean: Variance:	79 73 186 29 1.058	Packets late(discarded): Packets not found: Total packet lost: Total Packet loss %:	69 148 148 9.40	

15) For G729, 8 kbps, VAD is off, Jitter rate is %5, Packet loss rate is %5 and PLC is on,

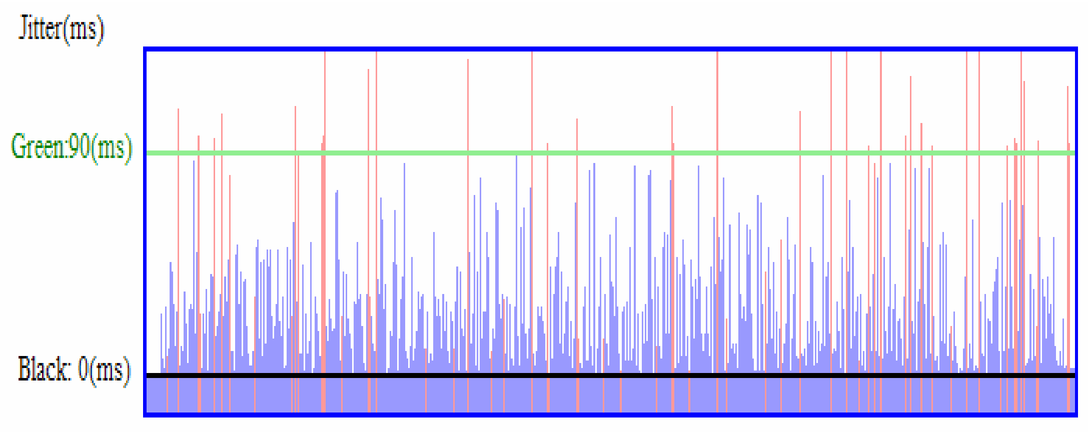


Figure 8.14 Test graph result

Table 8.11 Test results

Network Statistics						
Total Packets: 8 kbps Packets: Silence Packets: Empty Packets: Unknown Packets:	1575 1575 0 0 0	Packets lost: Jitter exceed 90 ms: Maximum: Mean: Variance:	78 92 202 32 1.231	Packets late(discarded): Packets not found: Total packet lost: Total Packet loss %:	83 161 161 10.22	

16) For G729, 8 kbps, VAD is on, Jitter rate is %5, Packet loss rate is %5 and PLC is off,

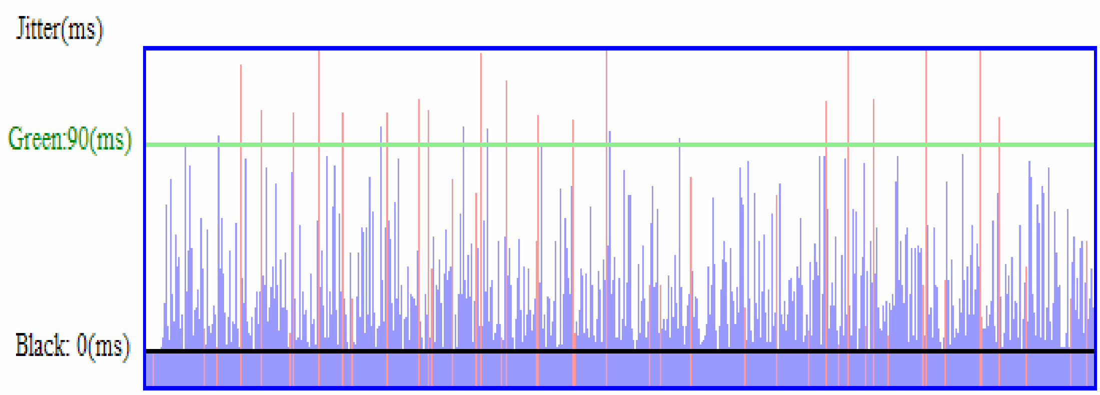


Figure 8.15 Test graph result

Table 8.12 Test results

Network Statistics						
Total Packets: 8 kbps Packets: Silence Packets: Empty Packets: Unknown Packets:	1575 1455 44 76 0	Packets lost: Jitter exceed 90 ms: Maximum: Mean: Variance:	79 69 191 30 1.075	Packets late(discarded): Packets not found: Total packet lost: Total Packet loss %:	50 129 129 8.19	

For all the tests, the differences about codecs, jitter and packet loss rates were observed. It is clear to understand from the tests that how jitter and packet loss are important for better voice quality.

8.2 MOS for tests

10 volunteers listened to these 16 audio files and an average mean opinion score was written to understand which codec has got better quality. Average results in the below,

Table 8.13 Mos ratios

Audio(*.wav)	Average MOS
1)	3.81
2)	3.63
3)	3.54
4)	3.32
5)	2.91
6)	2.89
7)	3.83
8)	3.75

Audio(*.wav)	Average MOS
9)	3.80
10)	3.71
11)	3.63
12)	3.37
13)	3.51
14)	3.49
15)	3.47
16)	3.23

Results have shown that the nearest scores to the original voice are 1 and 7.

At test 1; G723.1, 6.3 kbps codec , VAD is on, Jitter rate is %0, Packet loss rate is %0 and PLC is on.

At test 7: G726, 40 kbps, muLaw codec.

It is clear to understand from the all tests, if jitter and packet loss problems can be solved with some techniques like jitter buffering to the receiving end, packet loss concealment alghorithms or various packet recovery techniques, G723.1 6.3 kbps codec have enough quality for a conversation.

CHAPTER NINE
CONCLUSION

Voice over IP issue is growing up increasingly for a few years. Nevertheless, VoIP reliability and voice quality are keeping on to take important factors, specially when compared against the PSTN, that limit the extensive implementation of VoIP in end user markets.

In the VoIP consumer market, the whole things develops very quickly. Therefore, the obtainable bandwidth on networks will grow larger than now. This will also be helpful for the spreading of VoIP applications. When the available amount is adequately large, even high quality sound will be probable, which will definitely be a incentive for the use of VoIP programs. In addition, since compression methods are still developing, such high quality communications will be existing even sooner.

It is clear to understand that the greater delay or jitter in a VoIP structure, the greater packet loss. And the greater packet loss, the lower voice quality. Active monitoring and management of voice quality in a VoIP environment should assist recognize and lessen such unwanted incidences. As a result, voice quality is very important to be successful a VoIP system. Supervising VoIP can be quite onerous. Besides, a conversation quality section would assist monitoring packet loss, jitter, latency and voice quality scoring. And, an inventory section could discover and cope with the workings of the network.

REFERENCES

About.com, (2007). *Delay in VOIP*. Retrieved October 11, 2006 from http://voip.about.com/od/glossary/g/delay.htm

Cisco, (2007). *Understanding Delay in Packet Voice Networks.* Document ID: 5125 White paper. Retrieved February 07, 2007 from http://www.cisco.com/warp/public/788/voip/delay-details.html

Daigle, J. & J. Langford. (1996). Models for analysis of packet voice communications Systems. *IEEE Journal on Selected Areas in Communications. Vol. SAC-4, No. 6, pp. 847-55,*

Gammel, J. (1999). *Echo Cancellation for VoIP*. Retrieved December 17, 2006 from http://www.commsdesign.com/main/1999/10/9910feat4.htm

IEC. (2004). *Voice Quality (VQ) In Converging Telephony And IP Network.* Retrieved January 12, 2007 from http://www.iec.org/online/tutorials/acrobat/voice_qual.pdf

IEC. (2007). *Accelerating the Deployment of Voice over IP (VoIP) and Voice over ATM (VoATM)* Retrieved December 12, 2006 from http://www.iec.org/online/tutorials/voip_voatm/topic07.html

Kauffman, A. (2006). *An Algorithm to Evaluate the Echo Signal and the Voice Quality in VoIP Networks.* Ms. Thesis. University of Maryland In Electrical Engineering Department.

Ledford, J. (2006). *Cut the cord!. The Consumer's Guide To Voice Over IP.* (1th ed.). NY: Course Technology.

Liesenborgs, J. (2000). *Voice Over Internet Protocol In Networked Virtual Environments.* University of Maastricht, PhD thesis. Retrieved from january 13, 2007 From http://research.edm.uhasselt.be/jori/thesis/onlinethesis/contents.html

Montminy, C. (2000). *A Study of Speech Compression Algorithms for Voice Over IP.* B.A.Sc. University of Ottawa. AppIied Sciences in Computer Engineering http://www.collectionscanada.ca/obj/s4/f2/dsk1/tape4/PQDD_0017/MQ57147.pdf

Nortel Networks. (2001). *Voice Over Packet; An Assessment Of Voice Performance On Packet Networks.* White Paper. Retrieved November 23, 2006 from http://www.nortel.com/products/library/collateral/74007.25-09-01.pdf

Sinden, Richard. (2004). *Comparison Of Voice Over Internet Protocol With Circuit Switching Techniques.* White paper. Retrieved December 10, 2006 from http://mms.ecs.soton.ac.uk/mms2002/papers/28.pdf

Spirent. (2001). *Voice over IP(VOIP).* White paper. Retrieved December 14, 2006 from http://www.spirentcom.com/documents/100.pdf

Sun, L. (2004). *Speech Quality Prediction For Voice Over Internet Protocol Networks.* University of Plymouth, PhD thesis. Retrieved from november 24, 2006 From http://www.tech.plymouth.ac.uk/spmc/people/lfsun/publications/LSunPhDthesis.pdf

Sundstrom, K. (1999). *An Alghorithm To Evaluate the Echo Signal And The Voice Quality In VOIP Networks.* Voice Over IP: An Engineering Analysis. MSc Thesis, University of Anitoba, Department of Electrical and Computer Engineering.

Telchemy. (2006). *Impact of Delay in Voice over IP Services.* VoIP Performance Management Series. Retrieved 30 December, 2006 from http://www.telchemy.com/appnotes/Telchemy_Impact_of_Delay.pdf

Timothy, W. Kelly., (2002). *Voice Over İnternet Protocol (VOIP) for dummies.*(1th ed.). NJ: Wiley Publishing.

Tse, Lily. (2005). *Feasıbılıty Study Of Voıp Integratıon Into the Mysea Envıronment.* Ms. Thesis. University of California In Civilian, Naval Postgraduate School.

Voiptroubleshooter. (2004). *Indepth: Packet Loss Burstiness.* Retrieved November 03, 2006 from http://www.voiptroubleshooter.com/indepth/burstloss.html

Voiptroubleshooter. (2006). *Indepth: Jitter.* Retrieved December 09, 2006 from http://www.voiptroubleshooter.com/indepth/jittersources.html

Voiptroubleshooter. (2006). *Problem : Echo.* Retrieved December 13, 2006 from http://www.voiptroubleshooter.com/problems/echo.html

Wallace, Kevin. (2006). *Authorized Self-Study Guide Cisco Voice over IP (CVoice).* (1th ed.). Cisco Press. ISBN: 1-58705-262-8.

Wallingford, T. (2005). *Switching to Voip.* (1th ed.). NJ: O'Reilly.

Wikipedia. (2007). *Packet loss.* Retrieved December 16, 2006 from http://en.wikipedia.org/wiki/Packet_loss

William, C. Hardy., (2001). Voip *Service Quality; Measuring and Evaluating Packet Packet Switched Voice.* (1th ed.). NY: McGraw- Hill.

Woodard, Jason. (1995). *Digital Coding of Speech Using Code Excited Linear Prediction.* University of Southampton. PhD thesis. Retrieved December 09, 2006 from http://www-mobile.ecs.soton.ac.uk/speech_codecs/

Wright, David. (2001). *Voice over Packet Networks*. Wiley and Sons, Hoboken, NJ.