Emulation Study of Speech Communications over ATM Networks

by

Donglin Shen, B. A

A thesis submitted to
the Faculty of Graduate Studies and Research
in partial fulfillment of
the requirements for the degree of
Master of Application Science
Department of Electrical Engineering

University of Ottawa
Ottawa, Canada

ISBN 0-612-15672-9

UNIVERSITÉ D'OTTAWA
UNIVERSITY OF OTTAWA

Abstract

Speech communications over ATM networks is one of the important issues in the broadband ISDN. Although CCITT Study Group XIII has already proposed the Draft Recommendation I.121 on speech communications over broadband ISDN, there are many open issues to be further studied before the effective deployment of speech communications over broadband ISDN can take place.

In this thesis, several issues on speech transmission over ATM network have been studied, such as packetization delay, network queuing delay, digital speech encoding algorithms and PVR algorithm. A packetized speech emulation device is proposed to provide the capability of subjective speech transmission quality evaluation over ATM network or other kind of packetized network. The boundary of speech transmission quality degradation that human hearing can tolerate against information loss rate, delay fluctuation and encoding mechanism are found through emulation. The echo effect in ATM network, which is a primary issue in speech communications is also discussed in the thesis.

In particular, the emphasis is given to the specification, design and implementation of ATM speech emulator which consists of two subsystems: ATM Network Simulator (ATMNS) and Speech Transmission Emulator (STE). ATMNS has been specified according to the results of ATM network performance analysis, and STE is based on ATM specifications recommended by CCITT. A prototype of the emulator has been implemented on a personal computer and DSP5600 development system with a special designed audio interface to interconnect phone sets to the DSP5600 AD input. The software had been written in "C" and DSP assembly language.

Subjective evaluation are conducted in terms of following factors: Cell discarding rate, different network queuing delay and fluctuation to different PVR algorithms. These factors are basic issues which may affect ATM speech transmission quality in ATM network. Finally, test results under different network conditions are given.

Acknowledgments

I would like to thank my thesis supervisor Dr. L. Orozco-Barbosa for his help, encouragement over the course of my graduate study. I would like to particularly thank him for his many useful comments in the writing of the thesis.

I would like to thank my former thesis supervisor Dr. W. F. McGee for his help, encouragement and financial support over the course of my graduate study. I would like to particularly thank for his many useful comments during my work and studies in the University of Ottawa.

I would like to thank my former thesis supervisor Dr. N. D. Georganas for his help, encouragement and financial support over the course of my graduate study and this research project.

Table of Contents

List of Figures

List of Tables

Acronym Names

TIA	Telecommunication Industry Association	107
VLSI	Very Large Scale Integrated circuit	2
VTE	Voice Transport Emulator	62

CHAPTER 1 Introduction

In 1830, two weeks were expected for news to travel from Washington DC to New Orleans and three weeks were normal from New York to Chicago. But today, one hundred and sixty years later, we can call the people who live in the other side of the world just like to call our neighbor. If we use photonic network to provide news services, it only takes about few seconds for news to travel from an editor's terminal in New York to the reader's screen who may live in the Far East.

As technology advances, computers and telecommunications become even more interconnected, people will reach for information as a normal part of their every day lives - such as to make business decisions, treat illness, educate youngsters, conduct financial transactions, seek employment opportunities, and engage in hundred personal, business and entertainment activities. On the other hand, the increase of service requirements on the telecommunications market will stimulate the rapid evolution of telecom technology.

The breakthrough of computer, VLSI and fiber optic technologies have brought revolutionary changes in telecommunications and our daily life. The tremendous bandwidth provided by optic fiber makes it possible for one telecom network to meet all kind of services requirements, from narrow band data, voice services to wide band video and high speed data transfers. A most promising service that today's technology can support is teleconferencing which integrates data, voice and real time video services together into one application, so that the people around the world can engage in a conference over a telecommunication network just like in a meeting room. These are the objectives of the broadband service and the basic motivation of the deployment of the BISDN - Broadband Integrated Service Digital Networks.

The concept of ISDN first attracted people's attention in 1976, on the IEEE Communications Society ISDN Symposium[1]. The contentious issue was whether or not we needed digital switches in the local office. Unfortunately, general agreement was not reached until the fourth symposium in 1983. At this time, there was a heavy emphasis on obtaining the opinions of potential users as to their requirements for ISDN. Actually it was a real take off of ISDN. People started to think about the market of ISDN, how ISDN customer premises equipment should be developed and implemented.

The first intimation of broadband ISDN was on the 3rd ISDN symposium in 1981 by the keynote speaker - Irwin Dorros. He provided a vision of a large digital pipe to the premises. At that time, the idea of broadband ISDN was only a pipe, which did not cause much attention from attendees. It was not until the seventh symposium in 1989 that BISDN started to receive a good deal of attention for the first time. Since then, the number of research papers and publications on B-ISDN networks has just been growing at a tremendous speed.

Before ISDN was formally proposed, telecommunications networks were divided into two categories: connection-oriented circuit-switch networks which are mainly used for telephone services and connectionless oriented packet-switched networks which are mainly used for data transfer services. Packet switched network was originally designed for data transfer services, such as the ARPA net. In the last decade, a lot of research has been done on packet voice technology. But because of the variable packet delays commonly encountered in packet switch networks, the issue of voice quality has not been properly resolved.

1.1 ATM Overview

Asynchronous Transfer Mode (ATM) technology is expected to enable future Broadband Integrated Service Digital Networks (BISDN) to handle integrated traffic ranging from narrow band voice and data services to broadband high speed data transfer and video services, such as HDTV, video on demand and some other multimedia applications. The ATM approach provides a flexible means for the dynamic bandwidth allocation and multimedia communications to support a wide range of BISDN services.

The goal of BISDN is to define a user interface and a network that will meet these various service requirements. An objective of BISDN is to be able to accommodate dynamic changes in service mixes, both at the level of the individual interface as well as over the whole system. High-capacity and high-performance fiber-based transmission facilities are generally assumed to be required to support this environment. In addition, the network must have flexible switching capabilities and special-purpose service modules

"Asynchronous" has often been confused with asynchronous transmission. In fact, ATM is not just an asynchronous transmission technique. *Figure 1.1* shows the hierarchical relationships of some of the functional layers required for information transfer across BISDN.

The transmission layer, also referred to as the physical medium-dependent layer, underlies the transfer modes. A wide variety of service modules can be supported as higher-layer functions above the transfer mode layer. Connectionless services, multimedia, and digital signal processing capabilities can all be included in the BISDN as service modules. User-network signalling for control of connection-oriented and complex services also belongs at this level. A point-to-point connection-oriented service may not require a service module above the transfer mode layer, since ATM can be used to perform the relaying functions required to transport the user information through the network.

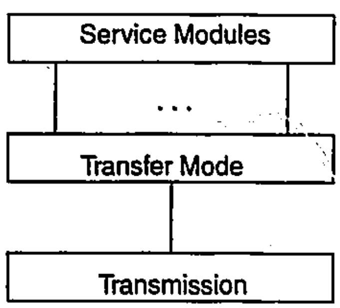


FIGURE 1.1 BISDN Functional Layer

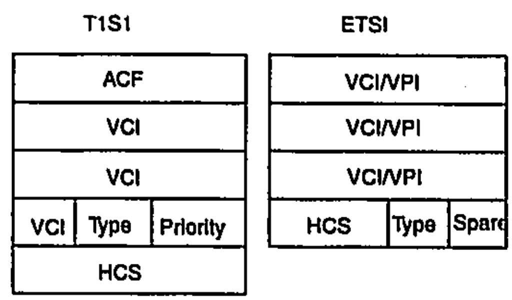


FIGURE 1.2 ATM Header Format

The choice of ATM as the transfer mode for BISDN is a fundamental shift in transmission principles. At the beginning, there were major disagreements on the most appropriate transfer mode for BISDN. The traditional Synchronous Transfer Mode (STM), based on time division switching and multiplexing, was initially assumed to be the appropriate transfer mode for BISDN by many, though there was a relatively small but vocal group of defenders for "new" ATM techniques.

The major issue that made ATM unpopular at that time was the degree to which voice service quality considerations should be a factor in determining ATM parameters. While it was recognized that voice services would have to be interworked with existing network facilities for a considerable period of time, views differed on the implications of this requirement on the specification of ATM parameters. Thus, the voice quality could be poorer relative to existing systems. There were also other factors to be taken into account in deciding these parameters. These could be more significant for the long-term utility of the network, e.g., suitability for high-quality video and high-speed data. These concerns and uncertainties of ATM networks do result in extensive research on the issues. These research efforts help to speed up the process of ATM network standardization. But the issue of ATM speech communication was still (and it is) not successfully resolved.

An ATM network is composed of cell processing nodes and digital transmission links. A cell is composed of a cell header (5 octets for North American proposed standard and 4 octets for European Standard, please refer to *Figure 1.2*)[1] for labeling and control, and an information field (48 octets North American and 32 octets was proposed by European) that carries the data block. An ATM node is composed of an ATM-MUX (multiplexer) including a coder-decoder (codec), a cell assembly/disassembly function, and/or cell switching and branching functions. In conventional circuit switching networks, speech is multiplexed every eight bits, and a

1. At the time this work was done, American and European groups have not reached the agreement on the cell size. The European community was still insisting on the 32 octets cell size for a better voice quality. At the time this thesis was written, both communities have already reached the agreement on a 48 octets cell sized as proposed by the North American community for a better network utilization. The design of the speech emulator had considered both proposals. In Chapter 6, the performances of both proposals have been evaluated as well.

synchronous frame signal is inserted into the time bit stream as a reference to identify a channel according to its time position. Each channel is in exclusive possession until the call is terminated, because of channel discrimination by the time position of the bit stream. Thus, the transmission rate of a channel must be constant. Conversely, in ATM networks, cells can be transmitted to the destination designated by the cell header only when information is being generated. All of the information required to label and control the cell is stored in the cell header. Therefore, the ATM is not tailored to a kind of information, it rather provides a flexible transmission mode for all kind of information. Because of this flexibility, the cell delivery is not perfectly guaranteed as in STM network when network congestion or other error condition occurs. These uncertainties will definitely have an impact to the fidelity of voice communications.

1.2 Objectives

The main objective of this thesis is to study, through emulation, the speech transmission quality issues in ATM networks. It includes the issues of cell loss rate, cell transmission delay and variation, PVR algorithm and echo impact.

The next section in this chapter will outline the structure of this thesis with a brief description of the contents of each chapter. Finally, the contributions of this study are highlighted.

1.3 Outline of the Thesis

In Chapter 2, after a brief introduction to digital speech communications technology, a review of existing problems of packetized voice and congestion control techniques in ATM networks is presented. The emphasis has been put on the impact of ATM network traffic parameters, such as cell delay and cell loss rate, to speech transmission quality.

In Chapter 3, we concentrate on one of the major problems encountered in packetized speech transmission - network queueing delay. Simulation models of ATM networks and cell delay distribution function are then discussed. Based on the characteristics of ATM cell arrival delay distribution, an approach for generating cell delays is proposed.

In Chapter 4, several packet voice receiving (PVR) strategies in regular packet switched networks have been reviewed and the strategies for ATM packet voice receivers have been developed based on the previous analysis. Emphasis has been made on how these PVR algorithms will compensate the delays that ATM cells experience in the network.

In Chapter 5, the design and the implementation of the ATM Speech Emulator is presented. The hardware and software configuration of the speech emulator are considered as well. The main elements of the real time emulation program are described therein.

In Chapter 6, a group of experiments is defined to verify the functionality of the speech emulator and some relevant issues in ATM speech communications. The results of the emulation have been evaluated using the Mean Opinion Score (MOS) system. The factors that influence the speech transport quality are analysed therein. Finally, experimental results are presented.

In Chapter 7, concluding remarks are given to evaluate this Speech Emulator implementation. Also further suggestions on ATM speech communications are given and discussed.

Finally, an appendix including the audio interface circuit is included for reference.

1.4 Contributions of the Study

This study has been presented in a TRIO Retreat, 1991 at Peterborouph, Ontario. It attracted a lot of people's attention to know what the speech emulator could do and how it was configured without special high speed equipment. In general, this study has achieved three contributions: 1) it is an experiment which proves that it is feasible to use low cost off-shelf hardware to build an ATM speech transmission emulator; 2) two ATM voice receiving strategies have been proposed and verified through emulation; 3) through subjective speech transmission evaluation, we have found out, under various network conditions, the limits of human hearing tolerance on ATM cell loss rate and cell transmission delay. It provides a useful reference to future ATM voice transmission study and ATM network design.

CHAPTER 2

Speech Communications and ATM Network

2.1 Speech Communications

2.1.1 Characteristics of Speech signal

When we talk over the phone, the voice is converted into an electrical signal which is transmitted through the telephone system to the listener. At the listener's end, this electrical signal is converted back to the audible voice, which makes the telephone conversation feasible. Based on this scenario, speech signal can be defined as an analog output signal of a conversion device which accordingly converts the acoustic energy generated from the talker into an electrical signal waveform.

In a telephone system, the telephone set serves as a transducer for converting acoustic speech energy to an electrical signal which can be transmitted along physical medium. When the telephone handset is in off-hook condition, a direct current is supplied by the telephone system over the wires leading to the customer's set. When we talk to the microphone in the phoneset, the changes in acoustic pressure will be modulated on the direct current at the audio rate. The ac component of this modulated current is commonly referred to as telephone speech signal.

The telephone speech signal at the central office has most of its energy concentrated in a band of frequencies from about 100 Hz to 5 KHz. This is a combined result of both the characteristics of the human voice and the band limitation introduced by a typical telephone set and loop [3]. This band is much wider than the need of the intelligibility, and its advantageous to further limit the bandwidth to improve performance in the presence of interference and noise. The optimum trade-off between economics and quality of transmission generally occurs when the telephone speech signal is band-limited to the range from about 200 to 3200 Hz with following properties[3]:

- frequency range: 200 to 3200 Hz
- signal-to-noise ratio: > 30 dB
- harmonic distortion: < 2 to 3%

2.1.2 Speech Signal Transmission

The basic speech signal is amplitude modulated at a syllabic rate. In addition, the speaker's pauses between phrases and sentences resulting in the speech energy being concentrated in "talk-spurts" of about 1-second average duration separated by gaps of a second or so. Thus the speech signal consists of randomly spaced bursts of energy of random duration.

Brady P.T.[53] has studied the conventional conversation activities between two talkers and formularized them into ten events. The four events which an average talker is aware of are the talkspurt, pause, double talk and mutual silence. They are defined as:

- Talkspurt: a continuous segment of speech between two silence period.
- Pause: a continuous segment of silence
- Double talk: a time when the speech is present from both talkers
- Mutual silence: a time when silence is present from both talkers

By referencing Brady's analysis, the speech signal in a telephone system can be summarized into two events - speech and silence. The speech is defined as a continuous segment of speech signal between two silence periods, which includes the near end talker's talkspurt and the far end talker's echo talkspurt signals if applicable. The echo is caused by the impedance mismatch of the 2 to 4 wires conversion hybrid in the telephone system. The silence can be defined as a continuous segment of no speech signal between two speech periods, which includes the pause and mutual silence. The speech signal transmission in the telephone system is

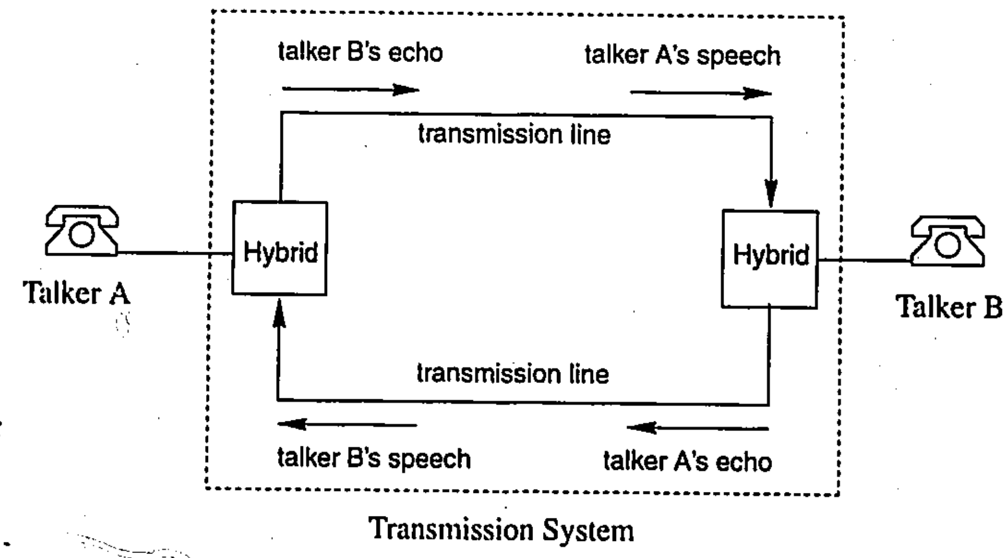


FIGURE 2.1 Speech signal transmission in telephone system

depicted in *Figure 2.1* . The speech signal on the upper transmission line consists of talker A's speech signal and talker B's echo signal which is reflected by talker A side's hybrid.

According to Brady's analysis, the means and standard deviations of the average of the four events are shown in *Table 2.1* .

Event	Mean (Sec)	Standard Deviation (Sec)
Talkspurt	1.197	0.444
Pause	1.846	0.648
Double Talk	0.251	0.055
Mutual Silence	0.466	0.088

TABLE 2.1. Mean and Standard Deviation of a Voice Conversation

In a normal conversation, the speech only takes about 40% of the total time, which indicates that 60% of speech duration is pause or silence. If the speech detection technique is applied to ATM network, it will make the speech signal transmission much more efficient, almost doubling the channel capacity.

2.1.3 Speech signal digitization

Since the interesting nature of the subject and its usefulness in a variety of applications, the field of voice digitization has received intensive study in the last 20 years[4]. These studies have produced many different types of voice digitizers (speech coder) with numerous variations of each type. The choice of a particular type is primarily dependent on the application and the level of voice quality desired. Generally speaking, applications may be categorized as: (1) transmission, (2) switching, (3) storage, or combinations of these[4]. Speech coders for transmission can be further classified according to wideband or narrowband applications. Wideband transmission applications involve T-carrier-like systems where standard telephone quality is desired. Speech coders used for ATM network belongs to this category. At present, economical telephone-quality voice digitizers require transmission rates between 32 and 64 kbps. Even greater data rates are required for digital transmission of program audio where higher quality is desired. For example.

384 kbps (including parity bits) has been proposed as a CCIR standard for sound broadcast signals [4].

Narrowband transmission applications generally arise when existing analog facilities are used such as the public telephone network or high frequency (HF) radio channels for secure voice applications. Voice digitizers used in digital cellular phone system are another typical application of the narrow band digitizer. These applications typically restrict the transmission rates to a range between 2 kbps and 16 kbps.

Since the internal bandwidth of a switching system is generally less constrained than the bandwidth of a transmission system, voice digitizers for switching applications typically use simple implementations with somewhat higher data rates.

In general, voice digitization techniques can be categorized into two classes, waveform reconstruction and analysis-synthesis. The first class encodes the analog waveform as faithfully as possible. It generates an output waveform which looks like the original input signal. This class is representative of the general problem of analog-to-digital and digital-to-analog conversions and is signal independent. Hence, it can code equally well a variety of signals such as speech, music and tones. The most commonly used method in the higher ranges of transmission rate is the Pulse Code Modulation (PCM) in which input speech is sampled in the order of 8 KHz and coded using 6-10 bits/sample yielding 48-80Kb/s digital signal. In addition to this standard method, several low bit-rate algorithms have been developed which take advantage of the correlation properties of the speech waveform [3]. The basic premise is that fewer bits are needed to code the derivative of the speech than to code the speech directly. Therefore, the difference between adjacent samples are coded for transmission instead of the individual sample as in PCM.

The second class is called analysis-synthesis which is used to produce a very low digital speech data rate for narrow band transmission systems or digital storage devices with limited capacity. The input speech is analyzed at fixed-duration intervals, frames, which are coded by a small number of bits for transmission. Some approaches for analyzing speech are the recognitions of speech phonemes, pitch periods and formats. A synthesizer can then reconstruct the speech by recreating the speech phonemes, pitch periods and formats based on the received coded data. Although this technique is capable of producing intelligible speech, generally the naturalness is lost.

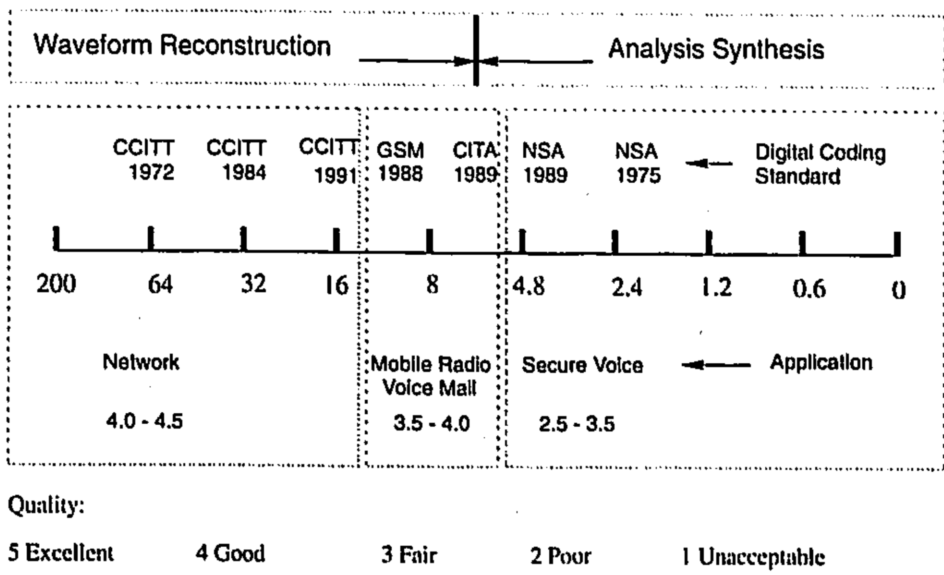


FIGURE 2.2 Spectrum of Speech Coding Transmission Rates

A spectrum of speech coding transmission rates of different coding standard are shown in *Figure 2.2* [6]. The figure highlights both waveform reconstruction coders that need relatively higher transmission rates and the analysis-synthesis coders for digitization at relatively lower bit rates. The figure also indicates the quality of speech reproduction that can presently be attained at a bit rate. The quality characterizations are denoted based on Most Opinion Score (MOS) system.[45]

Vocoders in the analysis-synthesis range provide a synthetic quality where the signal usually has lost substantial naturalness. Typically the signal sounds automaton like. Talker recognition is substantially degraded and vocoder performance is talker dependent.

Waveform codecs that provide communications quality speech can be realized with rates below 16 Kb/s. The signal is highly intelligible but has noticeable quality reduction, some detectable distortion and perhaps lessened talker recognition.

Telephone toll quality digital codecs can be realized for speech signals at coding rate between 16 Kb/s and 64 Kb/s. The term toll quality is typically used to imply quality comparable to that of an analog speech signal having approximately the following properties:

At the upper end of the spectrum, above 64 Kb/s, it is possible to obtain the signal-to-noise ratio and harmonic distortion characteristic of toll quality with input signal bandwidth significantly wider than normal telephone (e.g. 0 to 7 KHz or better). This grade of quality is referred to as commentary quality. It is appropriate for digitizing some varieties of radio broadcast material.

2.1.4 Echo in Speech Transmission

2.1.4.1 Echo

An echo, which is a reflection of a portion of the incident signal power, is produced in a transmission system wherever there is an impedance discontinuity. The most likely point of significant impedance discontinuity in the switched telephone network is at the connection to the customer loop in the end office.

An echo traveling in the opposite direction to the signal is often called a talker echo. If the talker echo is again reflected so that it travels in the same direction as the desired signal, it is called a listener echo. It has been observed that telephone circuits designed to limit talker echo usually provide adequate listener echo performance for voice transmission. The return loss on the data loops has been improved by the use of impedance-correcting networks.[3]

Talker echo, when audible, is at least an annoyance and, in severe cases, an actual impediment to the talker's normal speech process. The impairing effect of the echo depends on its amplitude and absolute delay. The amplitude of the echo is dependent on the circuit loss to the point of reflection, the return loss (commonly referred to as terminal balance) at the reflection, and the circuit loss in the transmission path back to the talker. The total acoustic-to-acoustic loss in the echo path is called the echo-path loss. Talker echo is controlled in the switched telephone network by inserting loss in the transmission path, by meeting minimum return loss requirements at points of impedance discontinuity, and by using echo suppressors or echo cancelers.

A reflection of the talker-echo signal at the talker's end of the connection may be heard by the listener and thus is referred to as listener echo. Methods used to control talker echo and singing generally control listener echo and therefore eliminate listener echo as a problem. Singing, or circuit oscillation, will occur if there is a net gain for energy circulating in the transmission path. A singing circuit is, of course, unusable and, in addition, can impair other circuits in the same facility. A near-singing condition, just before oscillation occurs, gives the circuit a highly objectionable hollow-sounding quality. Hence, adequate singing margin has to be provided; this is done through the control of circuit loss and return loss. The objective is to have at least 10-dB singing margin in 95 percent of all circuits.

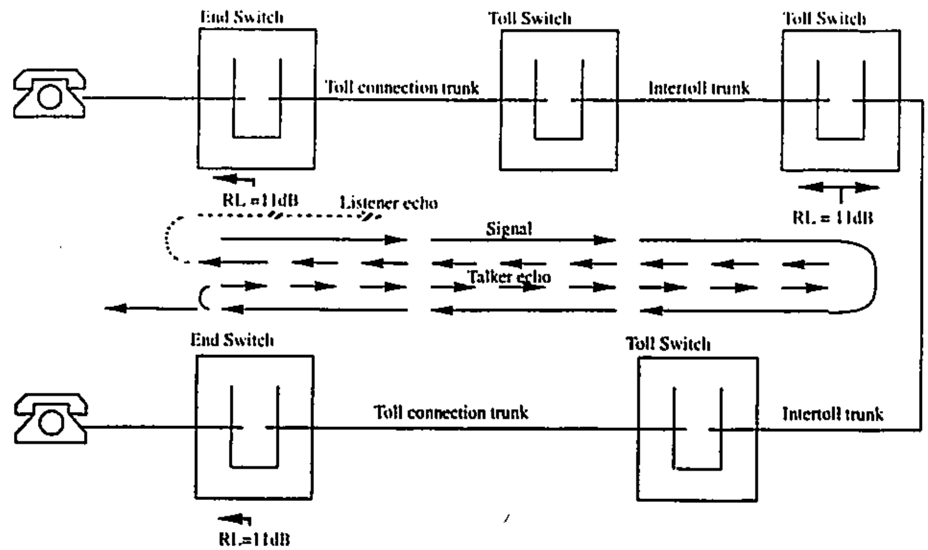


FIGURE 2.3 Return Loss and Echo Pass in Toll Connection

2.1.4.2 Echo return loss

The magnitude of the echo that a talker hears will depend on the echo-path loss, which is the sum of the return loss at the impedance discontinuity and the round-trip loss to that point in the circuit. As, previously mentioned, however, the customer's tolerance of the echo depends not only on the echo magnitude, but also on the round-trip delay between the echo and original signal. If the delay cannot be reduced and if return losses are improved as much as economically possible by impedance balancing, the echo magnitude can be decreased by increasing the electrical loss between the talker and the point where the mismatch occurs, at the unavoidable cost of reducing received volumes. Alternatively, the echo can be eliminated at the cost of installing echo suppressors or echo cancelers in the system.

Figure 2.4 shows how the grade-of-service (GOB) [3] opinion varies with echo-path loss for various round-trip echo-path delays. The curves show that the percent of GOB opinion decreases with decreasing echo-path loss and/or increasing echo-path delay. The least expensive means of controlling echo and meeting grade-of-service objectives is to introduce additional loss in the transmission path.

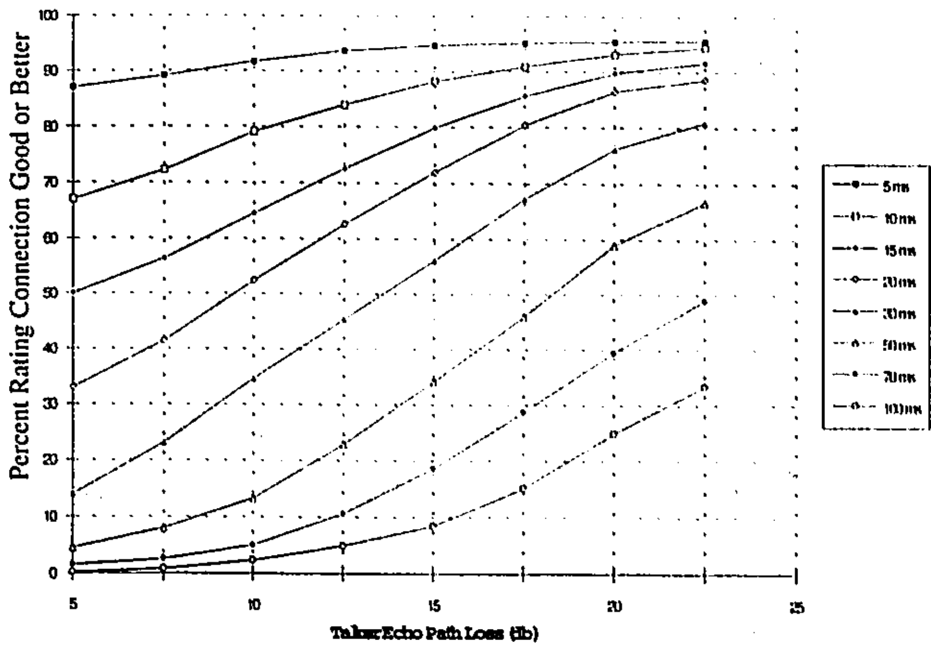


FIGURE 2.4 The Effect of Echo Path Loss and Delay vs. Transmission Quality

If the voice signal is carried in a purely digital format (e.g., via a digital carrier system), control of talker echo leads to a simpler trunk-loss administration plan. Fortunately, the delays encountered in a digital network are less than those in the analog plant because of lower delays in the multiplex terminals and digital-switching machines. Therefore, the optimum loss for echo is less than that required for the analog plant [3]. This as well as other considerations have made possible a fixed-loss trunk design for a purely digital-signal network. A loss of 6 dB for all length connections has been selected as the best compromise. This loss is obtained by allocating 3-dB loss to both the transmitting and receiving toll-connecting trunks which connect the analog voice-signal to the digital-multiplex terminals at the toll office[7].

2.1.4.3 Echo Suppressors - Echo Cancelers

An echo suppressor [8] is basically a pair of voice-operated switches which, while one party is talking, insert a loss of 35 dB or more in the echo return path. When both parties talk simultaneously, the talker whose signal is stronger at the echo suppressor point controls the switch. Sometimes, the echo suppressor is implemented as two split echo suppressors, one appearing near each end of the circuit. At each end, the loss is controlled only in the outgoing path, based on the signals received from both directions. Although they effectively suppress echoes, echo suppressors can introduce transmission impairments by sometimes clipping the beginning of words. The latest echo-suppressor design works on digital signals in a digital switch [9][10]. Time sharing and digital processing techniques permit improved characteristics of echo cancellation and also reduced costs.

Another technique for improving echo performance that is widely deployed in telecommunications network is known as echo cancellation[11]. This method uses signal correlation techniques to determine parameters of a filter that processes the incoming signal on the 4-wire side of a hybrid. The filter forms an estimate of the echo when an incoming signal is present. This estimate is subtracted from the signal on the return path.

2.1.5 Speech Quality Evaluation

After an analog speech signal is band-limited, digitized and passed through a digital communication network, a certain degree of quality degradation will be expected at the receiver. Obviously, for good voice communication, it is necessary to understand every word, to recognize the speaker and to be able to detect from the tone some other information that is not expressed verbally, such as the speaker's mood. There have been many studies about the components of the quality of speech communication but none of them is objective enough to automate the process and build a system for measuring the absolute objective quality of a sound reproduction system. This is because it is already rather difficult to recognize isolated words in an absolute objective way because recognition depends on the number of syllables, the choice of listeners, their conditions and experience, background noise and the particular speaker. However, there are many subjective techniques for measuring speech reproduction quality. These techniques are based on subjective evaluations by a test group of people indicating preference, intelligibility of isolated words and phrases, distinguishing rhymths and the like.

Measurement of the speech quality is a difficult and a long standing problem. The IEEE recommended three methods for subjective measurement based on individual preference [19]. They are the Isopreference Method, the Relative Preference Method and the Category-Judgement Method. These methods can only be used to evaluate speech signals which are generated by one-way communication systems. Systems, such as two-way systems in which interactive communication takes place, can not be evaluated by these methods.

For the purpose of the experiment, a commonly used subjective rating scale of 0-4 is used, whenever possible, to quantify the level of digital speech quality. The speech quality is assessed in terms of both subjective quality, defined using opinion equivalent Q (Q_{op}), and intelligibility [12]. Opinion rating is employed to assess satisfaction with speech quality. Five grades of speech quality are distinguished (excellent, good, fair, poor, and unsatisfactory) and weighted mean values are calculated using weights of 4-0. The value obtained is called the Mean Opinion Score (MOS). To normalize the sending and receiving quality factors, the equivalent noise method, applied to an opinion test, is used to assess speech coding. In this method, the test signal is compared with reference signals having variable speech-correlated noise levels. Opinion equivalent Q_{op} is defined as the speech to speech-correlated noise ratio of the reference signal recommended by the CCITT [13], whose MOS is equal to that of the test signal.

Intelligibility is assessed by articulation. Articulation is the ratio of vocal sounds sent from the sending end to those correctly received at the receiving end. The used for the articulation measurements consist of monosyllables. Sound articulation is defined as the ratio of sent phonemes to correctly received phonemes.

2.2 Speech Transport on An ATM network

ATM network is a high speed packetized network. In ATM, specific time slots are not assigned to a channel. Instead, the information is carried in cells, which consist of a header and an information field. The header contains a label that uniquely identifies a virtual channel and is used for multiplexing and routing. Current CCITT agreements specify 5 octets for the header and 48 octets for the information field size.

Speech signals are typically encoded using a 64 kb/s μ-law or A-law (PCM) encoding algorithm, where an 8-bit voice sample is generated every 125 μs. Packetization delay is defined as the delay introduced when packetizing multiple voice samples into an ATM cell at the input to an ATM network and then unpacking the cell into a continuous bit stream at the output. The packetization delay is made up of waiting time at the input and emission time at the output of the ATM network. As the voice samples are generated at rate of one per 125μs, the packetization delay associated with a putting n voice samples into the payload of an ATM cell is n x 125μs. Thus, for a 48-octet information field the packetization delay for 64 kb/

s voice is 6 ms and for 32 kb/s Adaptive Differential PCM (ADPCM) encoding algorithm is 12 ms.

There are two major factors in digital communications systems which degrade the digital speech signal transmission quality. One is information loss rate and the other is signal transmission delay. In circuit switched systems, since dedicated bandwidth is allocated before the actual communication is started, the information loss rate is very low (10^{-6} or better) which only occurs when channel transmission error takes place, and the signal transmission delay is constant. In a digital packetized telephone system, because of the packet delay variation caused by flexible channel bandwidth allocation, these two factors will cause more severe problems than in the circuit switched system.

2.2.1 Cell Loss Rate

Cell loss rate is defined as the percentage of ATM cells which are lost during the transmission process. In ATM network, cell loss occurs in conditions, such as network congestion, cell misdelivery, cell discarding or extensive cell propagation delay.

Network congestion

Network congestion could be caused by several reasons, bursty traffic source is one of the major reasons in ATM network to cause the congestion.

Most traffic source applied to ATM network are bursty, such as compressed video or high speed data. Such kind of bursty traffic source may generate cells at a near-peak rate within a very short period of time, and then just about second later, it may become completely idle contributing no traffic on the network. ATM network has to support a large number of such bursty traffic sources, each changing their traffic load dynamically. Thus, the network traffic can change rapidly, forcing the network to move from one degree of congestion to another even when the number of calls on a network is constant. Due to this, it is possible that a large number of cells are lost during congestion periods, even when the long term averaged value of cell loss rate is kept small. In voice communications, for instance, this bursty loss of voice cells may cause noticeable performance degradation, such as clicks, at a destination user.

Cell Misdelivery

The speech samples are packed in a 48 bytes ATM cell and the routing information is contained in the 5 octets header. If bit errors occur in the header field which can not be corrected by the error control code, the whole cell (48 samples) will be mis-delivered to other network node and eventually lost. In today's fiber optic network with very low bit error rate (10^{-9} or better), this is a very low probability event.

Cell Discarding

Cell discarding is the result of network congestion conditions. During network congestion, a congestion control mechanism will be activated. It will refuse the new connection request and also purposely discarding cells according to different algorithm in order to recover from the network congestion [15]. Cell discarding can also occur at ATM switch when the input or output buffer is overflow during the congestion.

Extensive Cell Propagation Delay

During the network congestion period, cells could be queued up at each switch node on the path, resulting in the cell propagation delay being increased. When the congestion reach certain level, the input/output buffers will be overflowed which results in cell discarded as described in previous paragraph. It is a trade-off between cell loss rate and cell propagation delay. In order to reduce the cell loss rate, a larger buffer has to be considered at the switch node to compensate the queuing delay. This will definitely reduce the cell loss rate. But it also increase the cell propagation delay which is a negative impact to the delay sensitive applications like voice and real time image [14].

2.2.2 Cell Transmission Delay

In the telephone network, normally, delay refers to the transmission time required for the talker's speech-signal to reach the listener. The subjective reaction to delay is very different from that caused by echo. Within the continental United States, the currently used terrestrial facilities may introduce up to about 20-ms one-way delay, depending on the length and type of facility. Overseas connections via submarine cable may have one-way delays ranging up to 100 ms[27]. Delays of these magnitudes have no effect on speech transmission. However, larger delays, of the order of hundreds of milliseconds, are within human reaction time and thus can be expected to affect the user's opinion of a circuit. Subjective tests have shown that there generally is no adverse reaction to one-way delays in the range of 300 to 600 ms (600 to 1200 ms round trip) by customers unfamiliar with delay and not anticipating it [3].

2.2.3 ATM Cell Delay and Variation(Jitter)

In ATM network, the end-to-end cell transmission delay consists of two components. One is the constant delay which is the delay incurred when cell is packetized and propagated through the network and other is variable network queuing delay. When different CODEC or packetization technique is applied, the packetization delay will be changed accordingly.

As we discussed in previous section, due to queueing, transfer delays experienced by successive cells of a given connection throughout the ATM network are different: this phenomenon is referred as Cell Delay Variation (CDV)[52].

Cell delay variation has direct impact on the quality of real time continuous signals transmission, such as real time speech or video communications. In order to overcome the problem caused by cell delay variation, cell playback compensation delay (See *Figure 2.5*) have to be considered as well. The playback compensation delay depends on the cell transmission delay and its fluctuation as well.

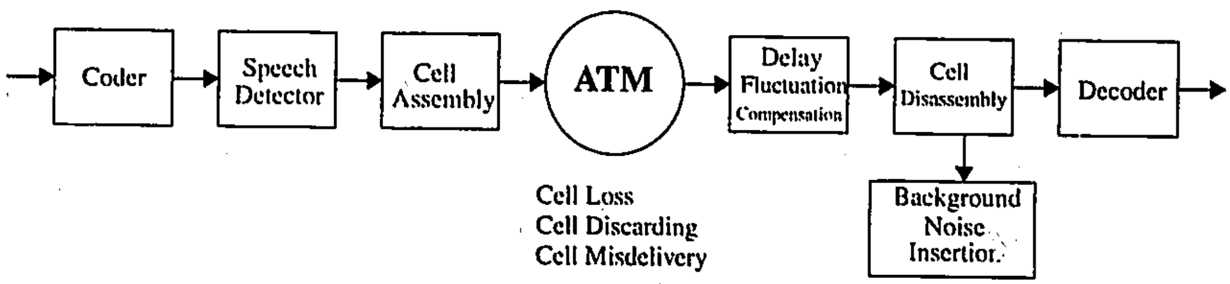


FIGURE 2.5 Speech Processing in ATM Network

2.2.4 Echo in ATM Network

There are two primary sources of echo existing in ATM network environment: echo arising from the customer interface to an ATM network (either the echo resulting from the acoustic path between the transmitter and receiver of a full-duplex telephone or the echo resulting from the use of a two wire telephone set via an ATM terminal adaptor), and echo arising from connections to the existing mixed analog/digital network.

Figure 2.6 illustrates these echo sources in ATM network connections. Section (a) identifies the acoustic echo generated in telephone set, (b) and (c) indicate the echo sources at the interconnection point with existing network.

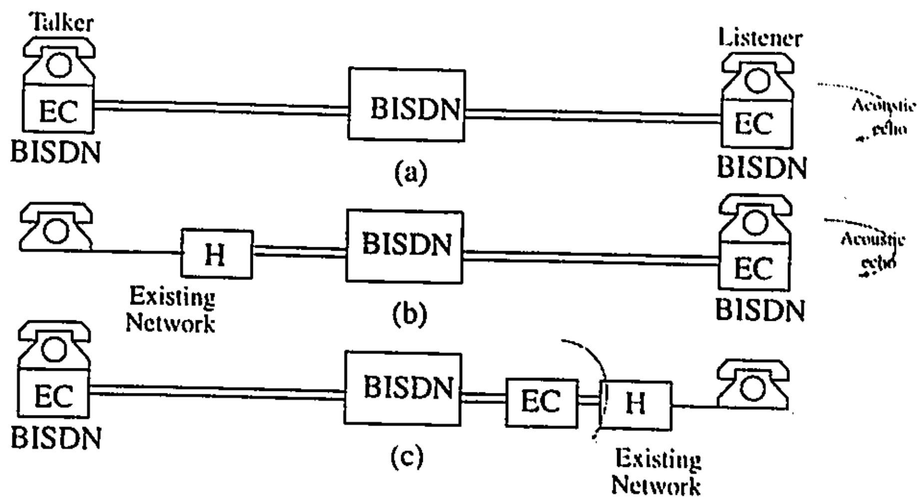


FIGURE 2.6 Echo Source in ATM Network

2.2.5 ATM Traffic Description Parameters

Traffic descriptors are a set of parameters which are used to accurately predict network traffic conditions and performance.

When a new connection is requested, the network needs to know the traffic characteristics of the new connection in order to accurately predict its ability to maintain a certain performance level. A set of traffic descriptors given from a user to a network will include sufficient parameters so that the network can accurately determine the user's traffic characteristics. However, from implementation point of view, a set of traffic descriptors should include the fewest possible parameters.

In general, the peak bit rate, the average bit rate and a measurement of burstiness are the most commonly used parameters as traffic descriptors. Among them, "burstiness" is the most important parameter, especially in an ATM network where most traffic sources are highly bursty. Burstiness is a parameter which describes how densely or sparsely cell arrival occur. It is well known that burstiness plays a critical role in determining network performance. At present, no consensus has been achieved as to how one can describe burstiness. Possible definitions of burstiness proposed include:

- the ratio of peak bit rate to average bit rate [16]
- the average burst length, i.e., the mean duration of time interval during which the traffic source transit at the peak rate [17]
- burst factor defined as the average number of bits cumulated in a buffer during a burst, which defined as (peak rate - average service bit rate) x average burst length [18]
- cell jitter ratio defined as the variance-to-mean ratio the cell interarrival times, which defined as, Var [cell interarrival times]/E[cell interarrival times[19]
- the squared coefficient of variation of the interarrival times, which defined as, Var[cell interarrival times]/E2[cell interarrival times[20]
- peakedness defined as the variance-to-mean ratio of number of busy servers in a fictitious infinite user group [21]

It is a difficult task to find a best approach to describe the burstiness more accurately and efficiently. This will definitely need further studies. Since the burst length will significantly affect the performance, it shall somehow be taken into account for the network traffic prediction. Normally, the longer burst length, the worse the network performance becomes, in the mean time, the cell loss probability becomes larger and the transmission delay becomes longer [17, 18]. On the other hand, the effect of the average burst length shows that with longer bursts, statistical multiplexing becomes less effects and thus fewer active sources can be supported for a given amount of bandwidth.

2.2.5.1 ATM Traffic Control Decision Criteria

Decision criteria is a group of network traffic parameters which are used to decide whether or not to accept a new connection. In ATM network, the cell transmission delays and the cell loss probabilities are the most commonly used decisions criteria in admission control, because they are good indications of the degree network congestion. When transmission delays and cell loss probabilities are applied in admission control, long-term-time-averaged values have usually been used [17]. Using a long-term-time-averaged value, however, it will not be sufficient in an ATM network because the network traffic can change rapidly and dynamically, forcing network to move from one degree of congestion to another.

Figure 2.7 on page 29 [22] shows how the cell loss probability changes in an ATM network as a function of time. In this figure the number of active calls jumps as network traffic varies, from *a* at time t_o, to *b* at time t_l, and to *c* at time t_2. At time t_3, the number of active calls decreases to *b*. The solid curve in the figure indicates the time-dependent behavior of the cell loss probability. For instance, when the number of active calls increase to *b* at time t_1, the network responds to the change and starts losing a large number of cells; gradually, the network goes up to the next level of congestion and reaches the value of the cell loss probability in steady state $P_{loss}(b)$. when another increase occurs at time t_2, the network responds again, gradually reaching the steady state, and so on. When the network traffic is highly bursty and changes dynamically, temporal network congestion can occur, and it is possible that a large number of cells are lost during congestion periods, even when the long-term-time-averaged value of loss rate is kept small. In voice communication for example, this burst loss of voice cells will definitely cause noticeable performance degradation (clicks) at a destination user. Therefore, some other decision criteria which takes temporal behavior of the network into account is needed.

Instantaneous cell loss probability [22] has also been proposed and used as a decision criterion to consider temporal behavior of a network. An instantaneous cell loss probability is a time-dependent cell loss probability (function of slot position or time), not the value averaged over a long period of time. The solid curve in *Figure 2.7* shows the instantaneous cell loss probability. The instantaneous cell loss probability is approximated by its steady state value and an approximate analysis is developed. A new connection is accepted only when the instantaneous cell loss rate is kept below a threshold value at each switching node for longer than a predetermined percentage of time.

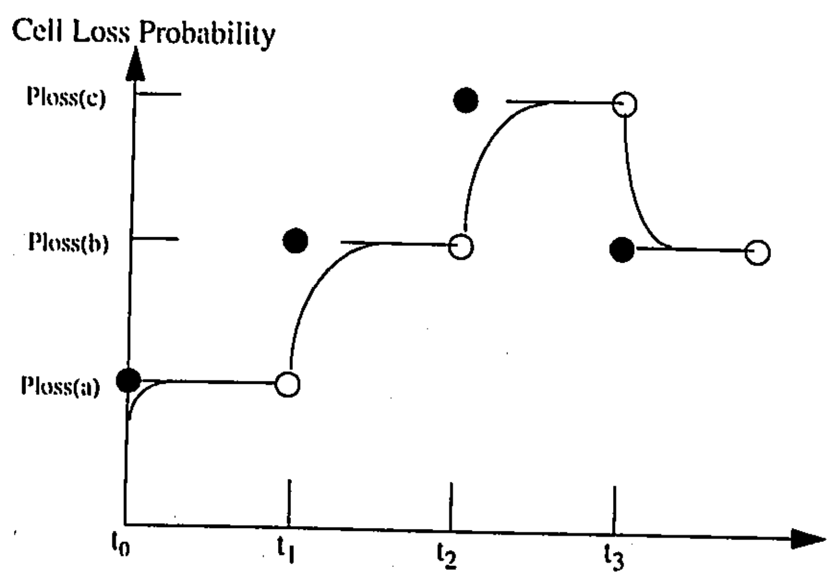


FIGURE 2.7 Time Dependant Behavior of Cell Loss Probability

The ineffectiveness of using the long-term-time averaged cell loss probability as a decision criterion has been demonstrated in [22] through numerical examples using realistic parameter values. It is shown that network congestion can last for a length of time on the order of a hundred milliseconds, even when the long-term-time-averaged cell loss probability is kept small. In voice conversation, this congestion period is comparable to a burst (talkspurt) length, and thus a whole talkspurt can be lost during congestion. It has also been shown that this burst cell loss can be avoided by using the instantaneous cell loss probability as a decision criterion in admission control.

2.2.5.2 Effects of Traffic Parameters on ATM Network Performance

To investigate the effect of various traffic parameters on network performance is one of the important research topics in admission controls.

In general, the network performance (the cell loss probability and the average delay) varies as a function of various parameters such as the number of sources, the peak bit rate, and the burstiness of the sources. Some of the common observations are.

- The average burst length is a very important parameter. As the average burst length increases, the performance degrades, i.e., the cell loss probability and delay time increase significantly [17,18].
- As the peak rate of each source is increased, the cell loss probability increases [17,18].
- In the case where homogeneous sources are multiplexed, if the offered load is kept constant, the cell loss probability decreases as the number of sources multiplexed increases. The reason for this is that when the number of sources multiplexed increases (keeping the offered load constant), the mean bit rate of each source decreases. The mean bit rate is a product of peak bit rate and the fraction of time in which a source is in the active state (i.e., the state in which a

source is transmitting at the peak rate). Therefore, the either the peak bit rate or the burst length (or both), the cell loss probability decreases in either case[18].

- In the case where heterogeneous sources are multiplexed, high-bit-rate sources dominate the performance; an increase in high-bit-rate traffic causes more significant increases in the cell loss probability than does an increase in low-bit-rate traffic. A similar observation is made in the case when homogeneous sources are multiplexed. Furthermore, when high-bit-rate sources are multiplexed, the fluctuation in the cell loss is larger than when low-bit-rate sources are multiplexed [22]. This is due to the fact that because of the high bit rate the number of traffic sources which can be multiplexed on one link is rather limited and not large enough to smooth out the bursty nature of each call.
- The cell loss probability decreases as the offered load decreases. Thus, a very efficient way to lower the cell loss probability is to decrease the offered load by providing larger bandwidth. This is only possible however, if one can assume that bandwidth is negligibly cheap.

According to these observations, we can conclude that the cell loss probability in ATM network is very sensitive to the network traffic conditions. In order to reduce the impact of the network congestion on the speech signal transport quality, higher priority should be assigned to the real time speech applications and the bandwidth allocated to voice communications should be as well guaranteed through effective network congestion control strategy.

CHAPTER 3

ATM Network Cell Delay and Loss Modeling

3.1 Cell Delay Distribution and Density Function

In order to emulate the impact of ATM network to the quality of speech signal transmission, it is mandatory to find out the cell delay distribution and delay density function in ATM network for building an ATM network simulator. As we discussed earlier, ATM is in many aspects similar to any other packet switching network. For voice transmission, it differs from a regular packet switch network in two aspects. First, ATM network uses very high speed transmission media. Normally, the transmission speed is 150 MBits and up. Second, ATM is a connection oriented network, the cells generated from one call will always propagate through one virtual path, which means that the cell sequence will never changed during the transmission.

3.1.1 Statement of the Problem

Since ATM packet voice receiver (PVR) has to act as a matching and recording device between the stochastic behavior of the network and the need for an ordered synchronous sample flow at the input of the D/A converter, the PVR design requires a description of external environmental influence on cells, which is summarized by the network delay random variable distributed according to the *pdf* $p(t)$. Since traffic flow is affected by $p(t)$ at the input of PVR, the objective of the study is to get a simple analytical expression for the overall delay *pdf* $w(t)$ and to design such a PVR for the best speech play-out strategy.

Voice transmission in ATM network takes place in the same way that in a packet switching network supporting connection oriented services. The main difference comes from the fact that the transmission speed in ATM network is much higher than regular packet switching network. The packet delay *pdf* derived on regular packet switching network can be taken as a reference for ATM cell delay *pdf*. Based on the experiment results, the Gamma distribution family will be taken as the reference network packet delay density function, because in this environment a good fit of the experiment delays [27] may be achieved with a simple expression for $p(t)$ (*EQ. 3 - 1*) and $P(t)$ (*EQ. 3 - 2*)

$$p(t) = \frac{\lambda^{k+1}}{k!} t^{k} e^{-\lambda t} u(t) \qquad \text{(EQ 3 - 1)}$$

and consequently

$$P(t) = \left(1 - \sum_{i=0}^{k} \frac{(\lambda t)^i}{i!} e^{-\lambda t}\right) u(t) \qquad \text{(EQ 3 - 2)}$$

The mean and the variance of this *pdf* are given, respectively, by

$$E(t) = \frac{k+1}{\lambda} \qquad \text{(EQ 3 - 3)}$$

$$Var(t) = \frac{k+1}{\lambda^2} \qquad \text{(EQ 3 - 4)}$$

In order to analyze the speech data packet end-to-end delay *pdf*, the statistical model of the speech packet delay can be stated as following:

1) Delay experienced by packets (cells) belonging to a single talkspurt are independently and identically distributed according to the network delay *pdf* $p(t)$ (in other words, the time spent by a packet in the network does not affect the delay of any other packet in the same talkspurt length).

In fact, this is not true when the system operates according to the "virtual circuit" routing scheme like in ATM network, which means that all the packets(cells) of the same call go along the same path. No packet/cell can overtake its previous packet/cell. Thus, its delay can not be less than the delay of preceding packet minus the packetization delay. It has been proved that the formula of the overall delay of the *pdf* remains unchanged when the reordering of talkspurt packets is made by the network routing policy or in the reconstruction buffer. Therefore, the assumed independence does not lessen the generality of the results [26].

2). Talkspurt durations are exponentially distributed according to $f(t) = \lambda_t e^{-\lambda_t t}$; this implies that the probability mass function of talkspurt length (in packets) is geometrical, i.e.,

Prob {talkspurt contains exactly (n+1) packets} $\equiv b_n = (1\text{-}\rho)\rho^n$

where $\rho = e^{-\lambda_t t_p}$; t_p = cell duration, $1/\lambda$ = mean talkspurt duration.

q_n = {prob actual cell is the $(n+1)$th of the talkspurt}= $q_n = \sum_{i=n}^{\infty} b_i/(i+1)$.

The talkspurt durations are distributed such that $q_n = (1 - \sigma)\sigma^n$, where σ is the parameter of the geometric probability mass function (*pmf*) that best fits the experimental observations.

3). Statistical independent between talkspurts is guaranteed (i.e., delays suffered by cells of every talkspurt are independent of those suffered by members of other talkspurts). This assumption is based on two reasons:

- The ratio between packet length and the capacity of channel and switching devices is so small that the state of the system queue change very rapidly.
- The speech decoder is assumed to recognize the long silence gaps in the voice waveform only. During a inter-talkspurt pause, the system processes many packet of other conversations and its state changes completely.

Please refer to reference [26] for detail explanation.

3.1.2 Average of End-to-End Delay *pdf*

Based on these assumptions and the network packet delay *pdf* discussed in *Section 3.1.1* , a general analytical expression of average end-to-end delay after PVR has been derived as following [26].

$$w(t) = \frac{1-\sigma}{1-\sigma P(t)} \cdot p(t-T) + \frac{\sigma(1-\sigma)p(t)}{[1-\sigma P(t)]^2} \cdot p(t)P(t-T) \qquad \text{(EQ 3 - 5)}$$

where: σ: parameter for the geometrical distributed probability mass function

$p(t)$: cell (packet) delay density function over network

P(t): cell (packet) delay distribution function

T: cell delay compensation

The packet delay *pdf* is given in *Figure 3.1*

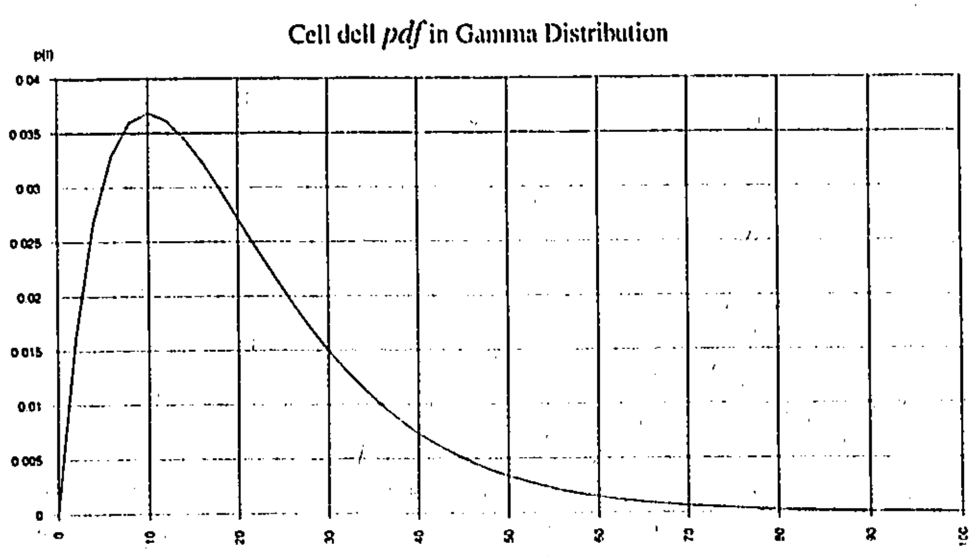


FIGURE 3.1. Probability density function of end-to-end delay after PVR

3.2 ATM Network Delay Simulation Model

Since ATM was proposed in 1988, the intensive research has been carried out on various aspects of the ATM network. Cell loss probability and cell propagation delay property are two of the major topics. Because of various architectures and congestion control mechanisms which have been proposed in ATM network, it is difficult to derive a universal cell delay *pdf(t)* for ATM network. Based on the literature study, most of the research on ATM cell propagation are based on simulations. These simulations normally rely on particular network structure and congestion control algorithm.

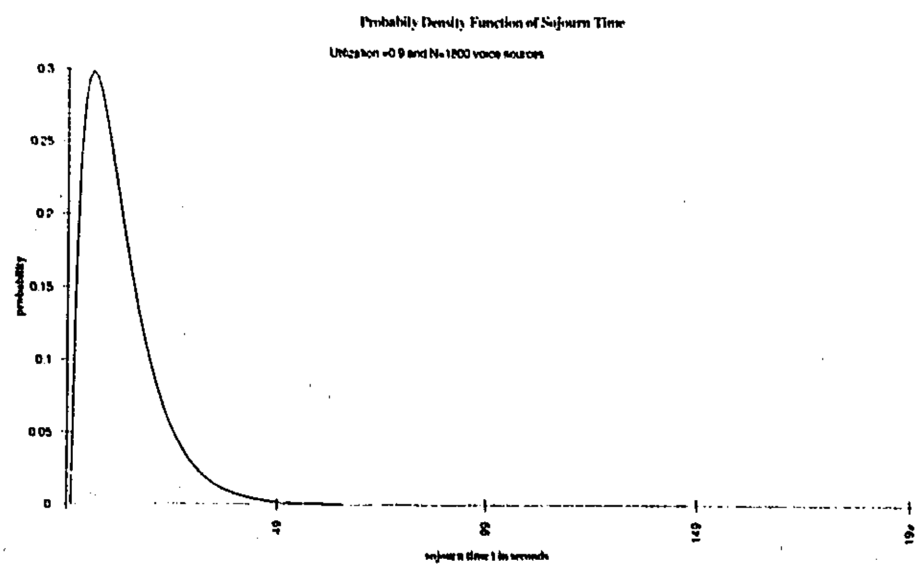


FIGURE 3.2 G. Ramamurthy's Simulation Result 1

In [29], Ramamurthy modeled an ATM type network delay by simulating a switch with stream traffic from K sources routed through an M x M space division packet switch. It is assumed that the switch employs output buffering with complete sharing. Each output queue has M input ports that are fed from the M output queues of the preceding stage. The simulation results are shown in *Figure 3.2* and *Figure 3.3* which are closely matching Gamma distribution.

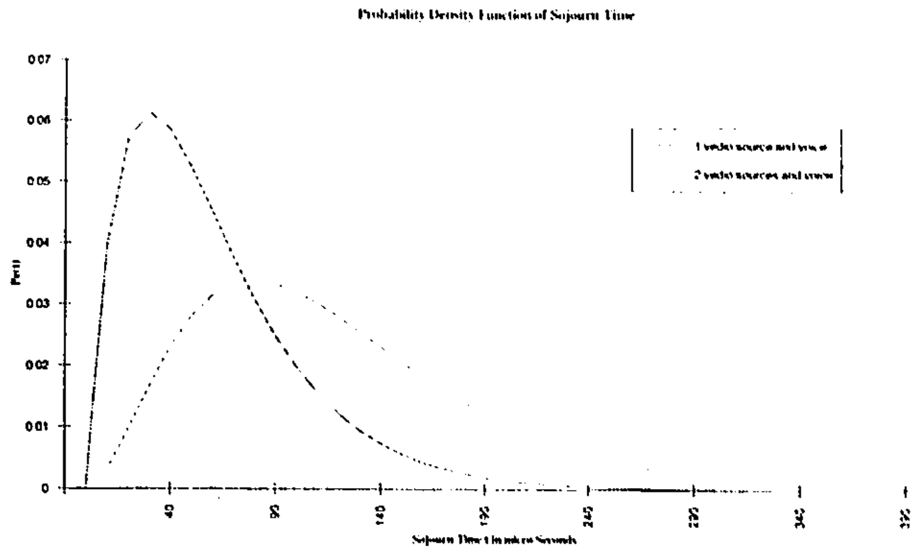


FIGURE 3.3 G. Ramamurthy's Simulation Result 2

In reference [28], authors have evaluated the maximum packet delay of a high speed packet network based on a "Bus Matrix" switch. According to the theoretical analysis and experimental result, the probability density function of "Bus Matrix" switch is obtained and given in *Figure 3.5* which closely matches with Gamma distribution as well. *Figure 3.4* is the simulation model of the switch.

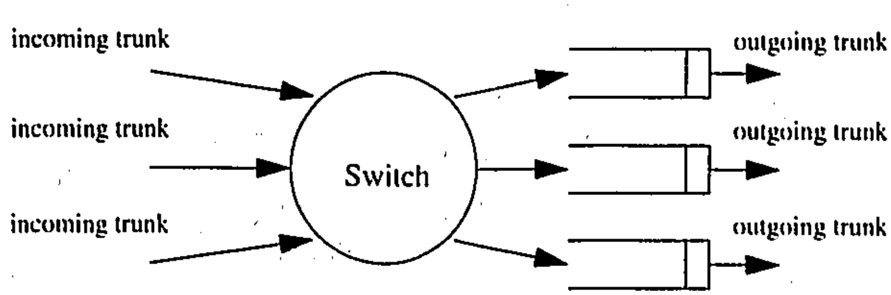


FIGURE 3.4 Ideal Packet Switch Model

Packet Delay Distribution

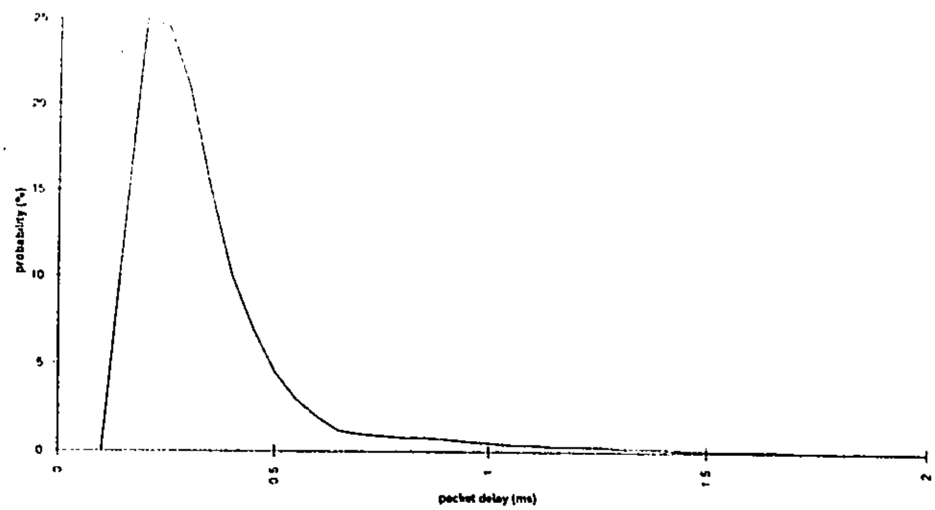


FIGURE 3.5 The Packet Delay Distribution

In reference [27], a probability density function (please refer to *Figure 3.6*) of the packet switched network is obtained through the actual measurement by J. Weinstein and W. Forgie. The measurement is based on a 10 hop paths through ARPANET. In this particular case, 99% packets experience delay between 200 and 700 ms. From the shape of curve, it more close to Weibull distribution.

Based on the examples which have been briefly mentioned above, following conclusions on probability density functions of packet switch network for speech signal transmission can be derived:

- Cell delay distribution in packet switch network is closely match with Gamma and Weibull distribution. For high speed packet switch network, the average delay is relatively short, in which the delay distribution lead to Gamma distribution and for regular packet switched network, the average delay is large (*Figure 3.6*), in which the delay distribution is more close to Weibull distribution.

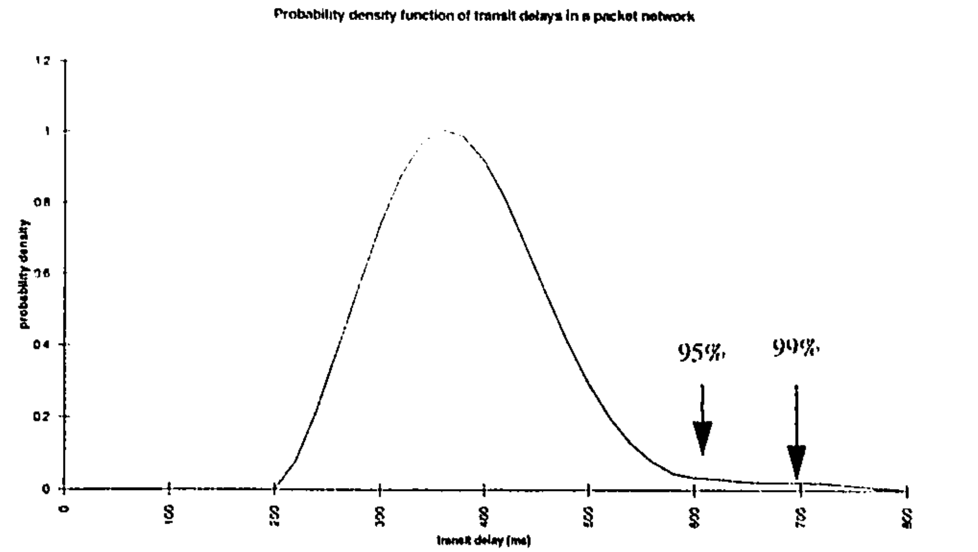


FIGURE 3.6 Probability Density Function of Transit Delay

- Delay variation caused by delay distribution exists and increases as average packet delay increases
- Because of the packet delay variation, a reconstruction delay (or called controlled delay) is required in order to play out the packets in the time interval as they are generated.
- Detail characteristic of cell delay is difficult to get and it is changing according to the network traffic. The distributions as mentioned in the first bullet is just a best approximation under particular experimental condition
- Since ATM is a high speed packet switched network, the Gamma distribution will be the closet fit for speech transmission in ATM network.

Based on these discussion, a Gamma distribution is selected to be used for the ATM network simulator. The proposed *pdf* is given in *EQ. 3 - 6*. Please refer to *EQ. 3 - 3* and *EQ. 3 - 4* for mean and variance calculation.

$$p(t) = \frac{\lambda^{k+1}}{k!}(t-\tau)^k e^{-\lambda(t-\tau)} u(t-\tau) \qquad \text{(EQ 3 - 6)}$$

where: τ: average delay shift parameter

$\lambda = 1/\beta$ and $k = \alpha - 1$

α is shape parameter and β is scale parameter

EQ. 3 - 6 will be directly used in the ATM speech transmission simulator to generate cell delay distribution which is used as the input of ATM PVR. One example is given in *Figure 3.7* .

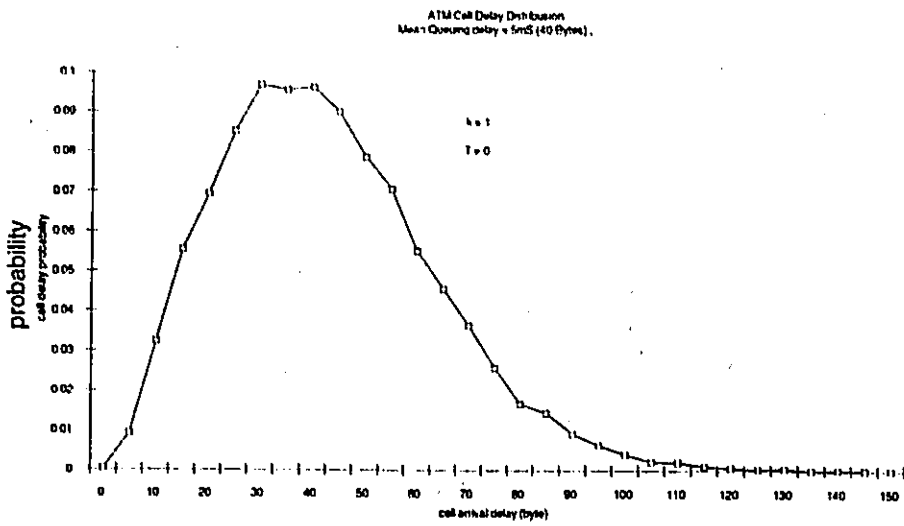


FIGURE 3.7 Cell Delay Distribution in ATM Speech Simulator

3.3 ATM Network Cell Loss Modeling

As stated in Chapter 2, cell loss in ATM network is mainly caused by the network congestion which is the result of bursty traffic in the network. A Poisson distributed cell loss property has been chosen to simulate the cell loss distribution in

ATM network, since it is a common practice in telecommunication network simulation. Poisson distribution *pdf* is given in *EQ. 3 - 7*.

$$p(t) = \frac{e^{-\lambda} \cdot \lambda^{t}}{t!}, \ldots t \in \{0, 1, 2 \ldots\} \qquad \text{(EQ 3 - 7)}$$

where: λ is mean that can be interpreted as average cell loss rate.

Figure 3.8 gives the actual cell loss interval distribution based on average cell loss rate of 0.1 percent. The data is generated from ATM network simulator, which qualitatively reflects the cell loss property in ATM network.

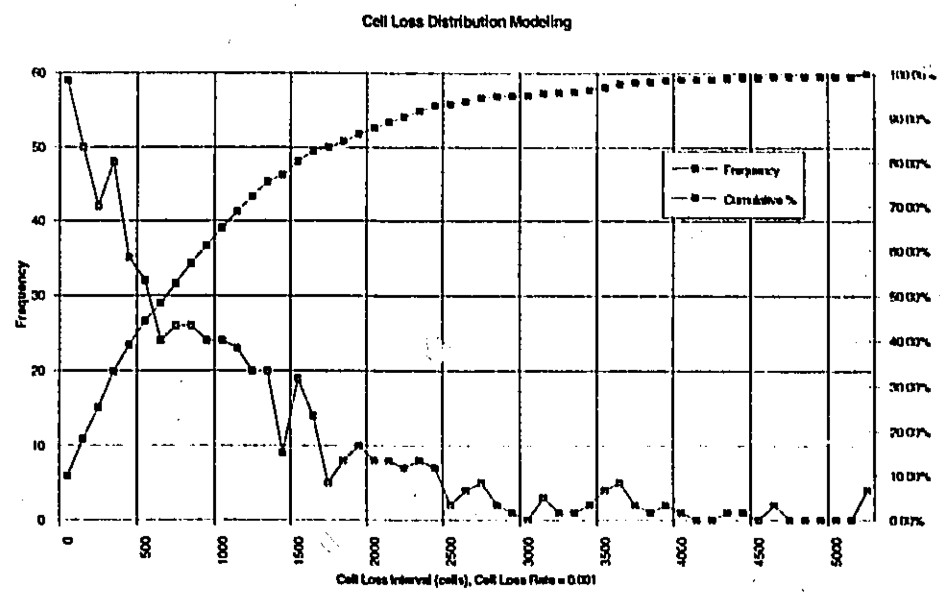


FIGURE 3.8 Cell Loss Distribution at Cell Loss Rate = 0.1%

3.4 Implementation Issues

3.4.1 Random number generation

In the ATM network simulator, two random number generators have been implemented to simulate ATM network traffic. The first one generates a sequence of numbers following a Gamma distribution used to simulate the cell delay variation in the network, measured in ms. The second one generates a sequence of numbers following a Poisson distribution and it is used to simulate the cell loss process.

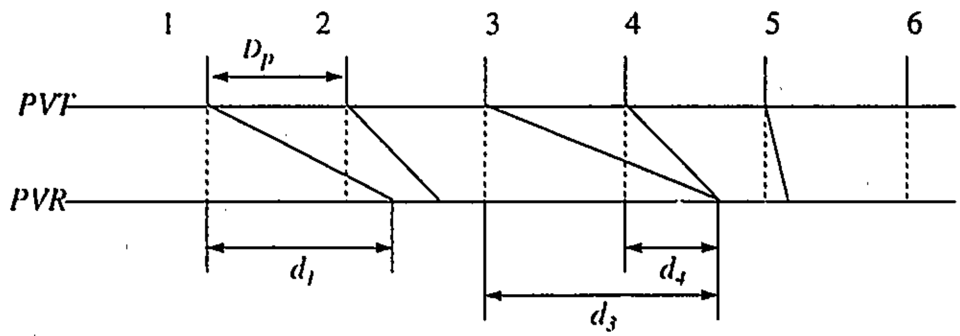


FIGURE 3.9 Cell Delay Relations in a Virtual Path

As discussed in *Section 3.1.1* , in an ATM network, the cells generated from one call are all transferred over a virtual circuit. This means that all the cells are transmitted through the same path, in sequence, and therefor a cell can not arrive at its destination before the cell preceding it. This characteristic imposes some constrains in the generation process of the sequence of random numbers representing the cell delay variation. If we assume that preceding random number representing the delay of previous cell is d_i and that the following random number in the sequence is d_{i-1}, then the d_{i-1} must be larger than (d_i - D_p), where D_p is packetization delay. If d_{i-1} does not meet this condition, d_{i-1} has to be rejected and another number must be generated. *Figure 3.9* depicts this condition through the relation of cell 3 and cell 4, which clearly indicates d_4 has to be large than d_3 minus D_p ($d_4 \geq d_3 - D_p$).

Because a certain random numbers are rejected, the shape of the delay distribution curve is slightly changed and the mean is increased in a small percentage as well. It is not an accurate Gamma distribution any longer. The actual cell delay distribution generated from cell delay simulator is given in *Figure 3.7* which is very close to the actual cell delay distribution derived from simulations.

The second generator is implemented in the ATM network simulator is the cell loss generator which generates Poisson distributed variables to simulate the cell loss caused by bit error, multiplexing error and some other error sources. The Poisson distribution *pdf* is given in *EQ. 3 - 7*. The input parameter λ is cell loss rate in percentage.

The random numbers generated are put into two separated data files. One is DELAY.DAT and another is ERROR.DAT. User should have the capability to decide the size of each file based the duration of the simulation. The data in the files can be repeatedly used during one simulation.

3.4.2 Control Parameters to ATM Network Simulator

Referring to *EQ. 3 - 6*, the λ, k and τ are user defined parameters which are used to control the ATM network simulator. λ and k are used to control the shape of the delay distribution density function. Actually it controls the cell delay variance of the network. τ is average cell delay shift parameter which shifts the *pdf* curve to the right on the diagram. Normally, the average network delay equals to $\left(\frac{k+1}{\lambda}+\tau\right)$. Because $\left(\frac{k+1}{\lambda}\right)$ is the Mean of the distribution $E(t)$ and τ is fixed network propagation delay, it means that no cell should experience the network transmission delay smaller than τ.

In order to obtain the proper k and λ, the $E(t)$ and $var(t)$ have to be correctly determined based on the simulation request, e.g average cell transmission delay and delay variance. According to *EQ. 3 - 3* and *EQ. 3 - 4*, λ and k can be obtained in following relations: *EQ. 3 - 8* and *EQ. 3 - 9*.

$$\lambda = \frac{E(t)}{Var(t)} \qquad \text{(EQ 3 - 8)}$$

$$k = \lambda \cdot E(t) - 1 \qquad \text{(EQ 3 - 9)}$$

where: $E(t)$ is average cell queueing delay in ATM network

$Var(t)$ is the queueing delay variance which reflects the jitter of cell arrivals

In this chapter, ATM network simulation models have been discussed. Based on the discussion, ATM cell delay and cell loss distribution models have been selected and evaluated. These models have been used in ATM speech emulator's design and deployment in Chapter 5 which are crucial to the success of ATM speech simulator.

CHAPTER 4

Packet Voice Receiver in ATM Network

4.1 PVR in Packet Switched Network

In a packetized voice communication network (*Figure 4.1*), speech is digitized at a uniform rate by the A/D encoder in the transmitting terminal, and then organized into packets by the packetizer. The speech detector judges each packet as to whether it contains active parts of the voice or not, and only non-silent packets are transmitted through the network on a store and forward basis. At the receiving terminal, voice packets are stored in the packet voice receiver, and then decoded into acoustic sound by the D/A decoder.

Since each voice packet waits at each intermediate node in the network until the outgoing channels become free, packets will arrive at the receiving terminal at random inter-arrival times. Because of fidelity requirement of voice, voice packet reassembly, which will play out voice packets at the same uniform rate as they were generated, is required at the packet voice receiver. [31].

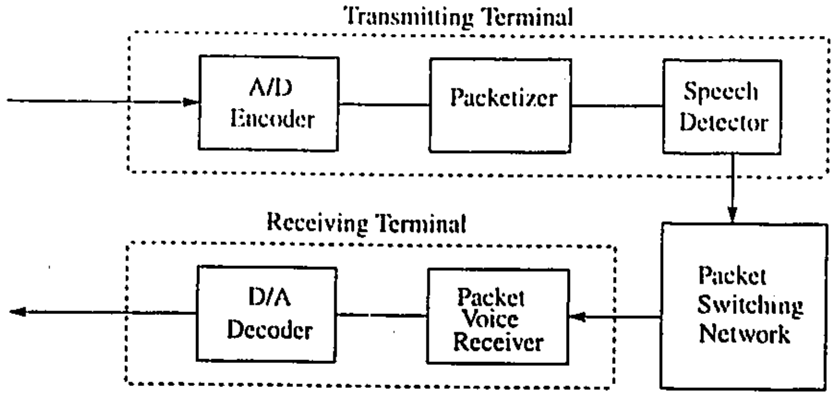


FIGURE 4.1 Packet Voice Communication Network

4.2 Performance Criteria for Packetized Voice Networks

Voice packet transmission delay may be one of the most important performance criteria for packetized voice networks. The voice packet total delay D is defined as time interval from the beginning of packetization to its played-out time. D becomes

$$D = D_p + D_q + D_t + D_r + R \qquad \text{(EQ 4 - 1)}$$

where

D_p: packet generation period

D_q: sum of queueing delays at intermediate nodes in a network

D_t: sum of voice packet transmission times in a network

R: propagation delay

D_r: depacketization delay (time interval from the arrival at the packet voice receiver to its play-out time

The transmission delay D_s of a voice packet can be further defined as the time interval from the beginning of its packetization to its arrival time at the packet voice receiver (please refer to *Figure 4.2*).

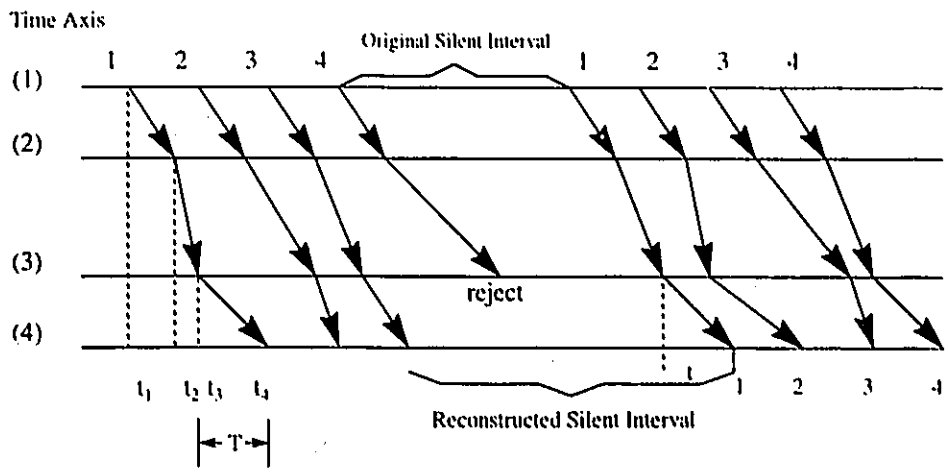


FIGURE 4.2 Null Timing Information (NTI) Strategy

Where:

Time Axis:

- t_1. Beginning time of packetization
- t_2. Beginning time of voice packet transmission
- t_3. Packet arrival time at the Packet Voice Receiver
- t_4. Beginning time of voice packet play-out

T: Control Time/speech reconstruction delay

$t_4 - t_1$: Packet total delay D *(beginning transmission to the start of play-out)*

$t_3 - t_1$: Packet transmission delay D_s

$t_2 - t_1$: Packet generation period D_p

$t_3 - t_2$: Queueing delay D_q + Packet transmission delay D_t + Propagation delay R

$t_4 - t_3$: Depacketization delay D_r.

Among these parameters, T is the most important one which has directly impact to the quality of voice signal synchronization. It is a major functionality for a packet voice receiver to properly determine the parameter T based on the network delay estimation.

4.3 Packet Voice Synchronization

One of the most significant technical problems in sending voice through a packet network involves the reconstruction of a continuous stream of voice from a set of packets (cells in ATM network) sent through the network[33].

As packets travel through the network, each packet can encounter a varying amount of queueing delay in the statistically multiplexed links. The variation in delay depends on the nature of the network, the traffic condition on the network and the speed of the network facilities. For a local area network, the variable delay is typically small, less than 10 ms [27]. For a long haul network, variable delay can be significantly larger, typically up to 100 ms.[27] The packet voice synchronization problem is generally more significant in a long-haul network than in a local network application of packet voice.

As the cells arrive at the packet voice receiver, they are reconstructed into a continuous stream of voice samples and delivered to the destination customer. Typically, this reconstruction is done by choosing a target play-out time for each incoming cell as a fixed interval, D_m, after the cell is generated[32]. Each cell that arrives before its target play-out time is placed in the proper sequence in a queue of cells from which the speech is reconstructed. If a cell does not arrive before its play-out time, the space of the cell in the queue will be empty, and the cell is effectively lost[1]. This is required to ensure that the entire system introduces a fixed delay into the speech path, and it makes that continuous speech signal be reconstructed without varying delay. The choice of the time interval should be made to optimize the delay and cell loss.

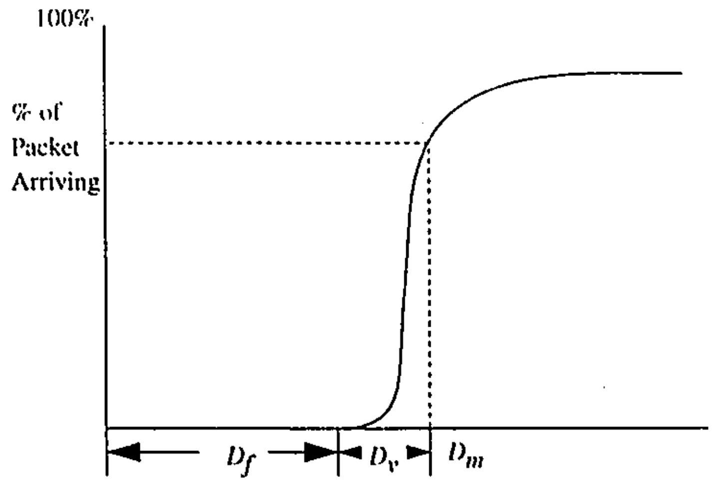


FIGURE 4.3 Delay Variation in a Packet Network

Figure 4.3 , which illustrates this trade-off in a typical packet network, shows qualitatively the proportion of cells that arrive in less than a given target transit delay for a particular connection. Fewer cells are lost if a longer delay is allowed because a larger proportion of cells will experience that delay, or less, in the network.

The delay experienced by a cell consists of a fixed delay (D_f), which is the same for each cell in a call, and a variable delay (D_v) which is different on each cell. The variable delay is the key factor and it introduces the trade-off between the delay experienced and the rate of cell loss illustrated in *Figure 4.3* . Wide delay variation in a particular network will result in a wide interval for D_v and more difficult

1. If next cell arrives before the previous cell play-out time, this cell can be play-out in the previous cell's slot.

packet voice synchronization. The nature of the variable delay is highly dependent on the nature of the network.

- In a local network, the variable delay is the result of contention for the network transport medium. The nature of the delay is highly dependent on the particular characteristics of the medium and the access protocols. For most proposed packet voice protocols, the variable delay is quite small.
- In a long-haul network where packets traverse multiple links and nodes, the variable delay results from queueing at each nodes and it depends on the number of links, utilization, and link speed.

The delay target mentioned earlier is limited at the upper end by the amount of delay that customer will tolerate in a call. While there is no hard limit to this delay, studies on customer acceptance of delay in calls [34], [35] suggest that 250 ms may be appropriate as an upper limit on one-way delay for a packet voice call in a public telephone network.

Once a delay target is chosen for a packet voice call, a mechanism is needed to determine the play-out time for each incoming cell. To achieve this goal, PVR has to take the first cell of a talkspurt arriving time t_a as the reference and add on the reconstruction delay $T \geq maximum\ D_v$ as the play-out time of the first cell (*Figure 4.2*). The following cells' play-out times will be determined on interval of packetization time, for example, k-th cell play-out time $T_{kp} = t_a + k\ D_p$, $k = 1, 2, \ldots n$, where the D_p is packetization delay. In an ATM network, D_p equals to 6 ms if 64 Kbits CODEC is used.

In order to keep the voice signal synchronization, PVR has to figure out the proper reconstruction delay T to compensate the delay variation caused by the network queueing. Network delay estimate is required for proper T determination. In fact, it is not an easy task to measure maximum D_v directly in a packet-switched network, because D_v is a part of the total cell delay and mixed with network transmission delay D_t together. Various methods can be used to estimate total transport delay of incoming cells, which can also be used to estimate D_v approximately. Among the most popular delay estimate methods being proposed are[32]:

- *Blind Delay:* the PRV makes a worst case assumption about the delay encountered by a cell.
- *Round trip estimation:* the roundtrip delay between PVS and PVR is used to estimate the one way delay of a particular cell
- *Absolute timing:* synchronized clocks are maintained at the PVS and PVR and the first cell of a talkspurt carries absolute time information.
- *Accumulated variable delay:* the network keeps track of the delay experienced by a cell as it travels through the network.

Among these methods, the simplest strategy for estimating the production time of an arriving cell is to make a worst case assumption - *Blind Delay*. Once the arrival time has been estimated, the PVR will determine the proper play-out time simply on each 6 ms interval started from the play-out time of the first cell of a talkspurt for each cell. This is the blind delay estimate, because the PVR makes its estimate blindly, with no information on the actual cell transit delay is required.

All other methods mentioned above require extra network functionality or network administration overhead to keep the timing information around. In an ATM network, the network timing information can not be transmitted in ATM cell header because of the constrain of header structure. If the timing information is required, it has to be transmitted in the cell body. This may reduce the cell payload significantly.

4.4 PVR in Packet Switched Network

According to the voice synchronization techniques and delay estimate methods, the following packet voice receiver design strategies have been proposed [32] for packet voice receiver design and development.

4.4.1 Null Timing Information (NTI) Strategy

Null Timing Information Strategy is illustrated in *Figure 4.2* . The packet voice receiver delays every first packet of the talkspurt by a given amount T of time (control time/reconstruction delay) and plays out succeeding packets at the same uniform rate as they were generated. If a packet is not received by its played-out time, that packet is considered to be lost. This strategy, requiring no network synchronization, is easy to realize; however, overall packet transmission delay may be relatively large and fidelity of played-out silence intervals may be low.

4.4.2 Complete Timing Information (CTI) Strategy

The Complete Timing Information Strategy is illustrated in *Figure 4.4* . If the network delay of a packet is less than a given control time T, that packet is additionally delayed at the receiver by an amount equal to the control time T minus its network delay, and then is played out. A packet with a delay greater than T is considered to be lost, even if it is the first one of the talkspurt. This strategy requires the network to be synchronized, and also requires timing information in the packet header. However, unlike the NTI strategy, this will keep overall packet transmis-

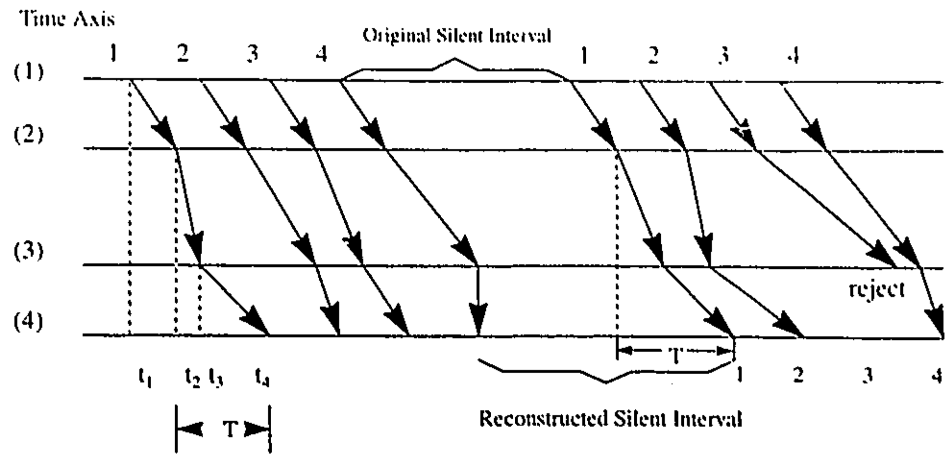


FIGURE 4.4 Complete Timing Information (CTI) Strategy

sion delay less than some constant, and will ensure relatively high fidelity of played-out silence intervals.

4.4.3 NTI-CTI Mix Strategy

If the network delay of the first packet of the talkspurt is less than a given control time T, then that packet and successive ones are played out in the same way as that in the CTI strategy. If the network delay of the first talkspurt packet is greater than T, the packet voice receiver plays out that packet immediately upon receiving it, and continues to play out successive packets at the same uniform rate as they were generated. Packets which are not received by their played-out times are considered to be lost.

4.5 Packet Voice Receiver in ATM network

Since ATM is a very high speed packet network and the network delay variation, comparing to regular packet-switched network, is relatively small, a simple and effective delay estimation method and PVR strategies should be used to achieve higher efficiency. In this section, we will concentrate on ATM packet receiver and delay compensation strategies.

In general, the PVR designed for regular packet network can be easily transferred to ATM network with some modifications. Based on the PVR synchronization strategies used in regular packet switched network and the properties of ATM network traffic conditions, two ATM PVRs are proposed and evaluated for the applications in ATM speech emulator design.

4.5.1 Fixed Delay Compensation PVR

This ATM PVR which is depicted in *Figure 4.5* is similar to the NTI PVR described *Section 4.4.1* . The worst case assumption for the PVR assumes that the cell on which the estimate is based arrives with minimum fixed transit delay (D_f only), and that other cells may be delayed by significantly more time. Hence, the PVR must set the target play-out time for the first cell to be its arrival time plus a maximum variable delay (Max D_v). Max D_v is the difference between the delay target (D_m) and the fixed network delays (D_f). Max D_v is chosen as reconstruction delay T so that cells that experience more than this delay will be lost in the play-out procedure without unacceptable degradation of speech quality. An additional small delay is added to the target play-out time to allow for clock drift between the PVS and the PVR since relative timing will be used for subsequent cells.

The delay estimate can be revised at any time since a new estimate can be made on any cell. A new estimate may differ from the current estimate by at most Max D_v. The delay target can be adjusted based on a new estimate by extending or contracting a silent interval.

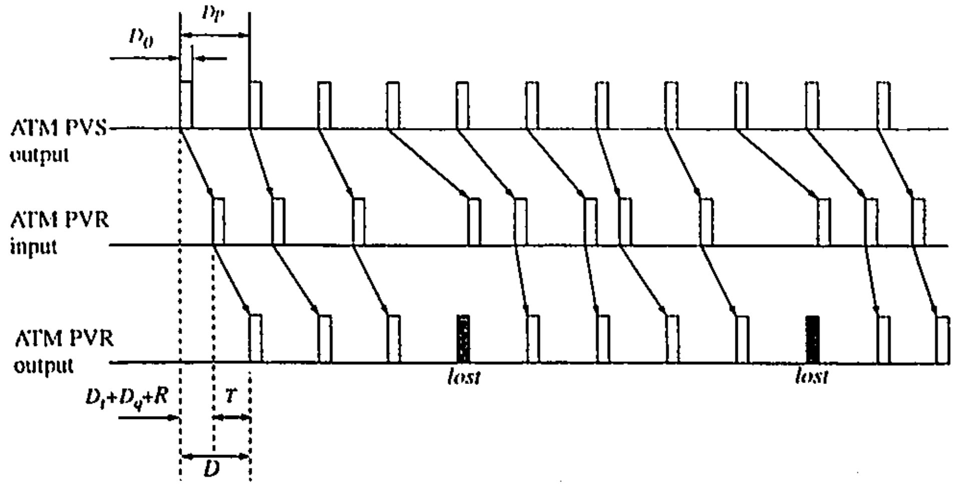


FIGURE 4.5 Fixed delay compensation ATM PVR

This method is very simple and requires no additions to the network to aid in packet voice reconstruction. For a local network application, D_v and Max D_v is relatively small, and other fixed delays are small as well. Blind delay will be very appropriate strategy with the estimate being made either on the first cell of the call only, or on the first cell of each spurt of speech. The adjustment of silent intervals to update the delay target should not pose a problem because the amount of adjustment will be small. For a long-haul network, however, Max D_v may vary in a wide range depending on the network structure and the traffic conditions.

4.5.2 Adaptive Delay Compensation PVR

Since the network traffic condition can change in a wide dynamic range, from time to time, it requires that extra delay be introduced in the call to avoid packet loss when the delay estimate is in error. Blind delay and round-trip delay estimation may produce estimates with substantial errors, while absolute timing and added variable delay are subject to small errors due to clock drift and propagation delay variation. These techniques can be improved by making use of additional information to adapt the delay estimate as the call progresses.

The adaptation can be made to change the target delay during silent intervals, so that the silence is compressed or expanded by a small amount. Compression or expansion of silence by small amounts is not noticeable in the reconstructed speech.

The adaptive change in play-out can be based either on the amount of speech in the buffer (i.e., the length of time between cell arrivals and target play-out time), or on the basis of repeated round-trip delay measurements. In either case, the repeated measurements given an indication of whether the current estimate is too high or too low, and the play-out targets are adjusted accordingly. The variance of inter-cells arrival time in each talkspurt will be used to estimate reconstruction delay adjustment. The relationship between delay variance and the mean of delay is described in *EQ. 4 - 2*, which is derived from *EQ. 3 - 3* and *EQ. 3 - 4 on page 34.*

$$E(t) = var(t) \cdot \lambda \qquad \text{(EQ 4 - 2)}$$

Adaptive approaches can also be of benefit with absolute timing and with added variable delay in determining the maximum expected transit delay. Most of the time the network will operate with considerably less variable delay than that experienced in the worst case, and transit delays can be reduced by reducing buffering in the PVR without losing any cells.

- Adapting play-out delay only on the basis of the number of cells arriving too late for play-out is quite difficult because few lost cells can be tolerated to give acceptable performance. Thus, the occurrence of lost cells should be infrequent with respect to changes in delay in the packet network, and it may be quite difficult to adjust playout delay based on lost cell rate and still achieve an acceptable lost rate.

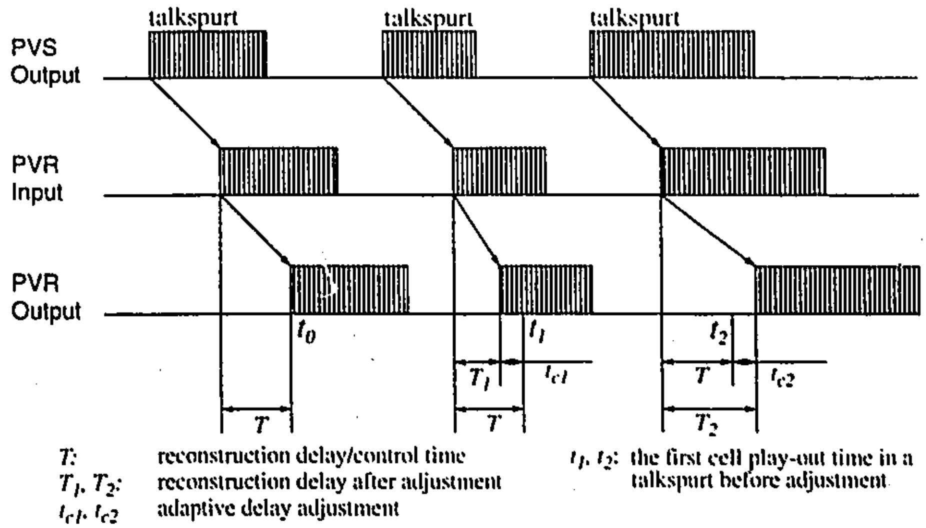


FIGURE 4.6 Adaptive Delay Compensation PVR

- Network conditions can change rapidly during intervals where no cells are sent. Thus a conservative delay estimate should be used after a prolonged silent period.

The various techniques considered in this section provide the basis for reconstructing speech from a stream of cells. The simpler techniques appear to be quite adequate for a local network application of ATM voice, while more complex techniques - Adaptive Delay Compensation may be needed to achieve adequate performance in a long-haul network with larger variable delay.

These two PVR algorithms which are discussed in the section were proposed for ATM voice communications by this thesis. They will be analyzed and evaluated through the ATM speech emulator design and subjective speech transmission assessment in following chapters.

CHAPTER 5

Design of an ATM Speech Emulator

5.1 Objective to Build Speech Emulator

5.1.1 Motivations

As discussed in earlier chapters, the end-to-end cell transport delay, cell delay variation and cell loss rate caused by the network traffic conditions are the major factors which may seriously degrade the service quality of the system. That is why these issues have to be studied carefully to gain better understanding of the relationship among them. For example, when the system interworking with existing 2-wire lines, the round trip delay has to be kept within the boundaries to avoid use of echo cancellation, where these limits are not clear at present and should preferably be studied and tested[36]. Another important issue to be studied is the one way delay. For long haul or international calls, some studies have shown that the quality of ATM voice communication is not sufficient [39]. Whether that is only caused by echo, echo processing or other reasons is not clear. More studies need to be done to find out the maximum acceptable delay value for ATM voice communications. A second issue is the maximum end-to-end delay that the PVR algorithm can compensate, which is really critical to voice communications.

To build an ATM speech transport emulator is the object of this study. Through emulation we can find more accurate answer to these questions than what the analytic approaches can provide to us. Many experiments can also be done on the emulator to investigate the interactions of these issues and some network parameters. One of most attractive features of speech emulator is the capability to perform subjective speech quality evaluation on emulated ATM network. The speech emulator can operate in real time to sample the speech signal, assemble ATM cells, add transmission errors and recover the speech signal after some of cells are delayed, lost or discarded.

5.1.2 Applications

Because ATM is a high speed packet switched network, all the issues in packet switched voice network will exist in ATM network as well. How will these issues affect voice transport quality in ATM network? What kind of end-to-end cell transmission delay and cell loss rate customer can tolerate in ATM voice transport? How does ATM network structure and congestion control algorithm affect the voice service quality? etc. ... All theses questions need to be answered before ATM is adapted to the public network. To reach the goal, ATM speech emulator will play an important role in ATM speech communications studies.

Currently, CCITT AAL1 protocol has defined that if the speech data cell can not arrive the receiving node within a certain delay limit, the cell will be discarded at the receiving node. How should this delay threshold be related to the network traffic condition? If a call has to go through multiple transport nodes, how should this delay threshold be set in each node to compensate the extra cell losses caused by excessive cell delay? What is the best trade-off between cell loss rate and cell delay limit when ATM network is designed and implemented? ATM speech emulator is aiming to answer these questions as much as possible.

Under certain network conditions, the emulator can also be used to test the service quality variations caused by:

- different receiving algorithms
- different speech encoding algorithm
- speech/silence detection algorithm
- loss cell recovery algorithm
- the maximum acceptable network delay at different traffic conditions that user can tolerate.

Here are three examples to explain how the speech emulator can help on the ATM speech transportation study.

5.1.2.1 Reconstruction Delay Versus Cell Delay Variation

Reconstruction delay is used to compensate variable delay of cell arrivals. When the cell delay variation increases, in order to keep the same cell loss rate, the reconstruction delay has to be increased. When reconstruction delay is increased, the total speech transport delay increases as well. How to arrange the reconstruction delay versus cell delay variation to keep the service quality an acceptable level?

5.1.2.2 Relationship among Cell Delay, Cell Loss and Echo Effect

When total cell delay increases, the chance of echo interference will increase as well. If we fix the echo pass loss, when the total cell delay increases to certain point, the echo interference will definitely become intolerable. Through emulation, we can find out the relationship between total cell delay and the echo interference at different echo pass loss.

5.1.2.3 PVR Algorithm Comparison and Evaluation

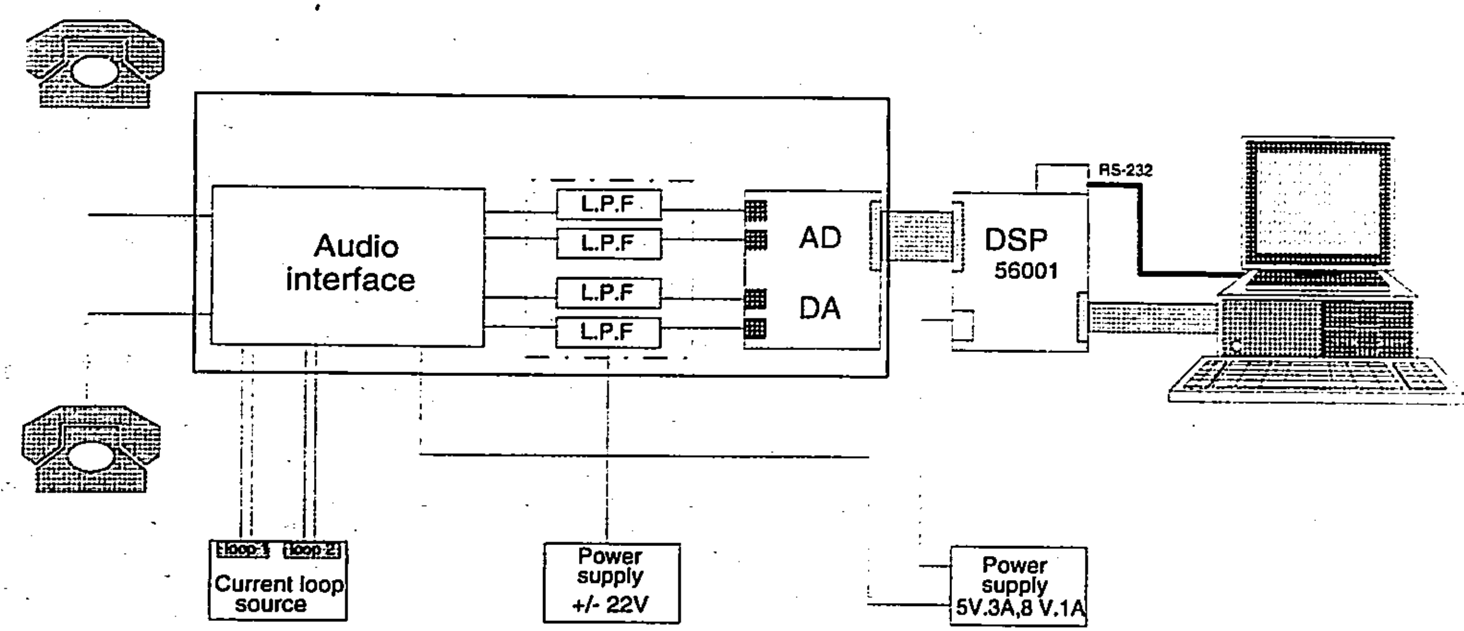
The packet voice receiver algorithm is a key component in ATM voice communication. Based on different synchronization strategies and timing estimation methods, there are many different packet voice receiver algorithms available in this field. Through an emulation study, we can evaluate the service quality and performance of different PVRs and different PVR algorithms.

5.2 System Architecture

The structure of the ATM speech transport emulator is depicted in *Figure 5.1* . It consists of three major components: a) ATM network simulator which simulates the ATM network behavior on the end-to-end cell transmission delay and cell loss rate basis. It is implemented on a PC with menu driven user interface; b) ATM voice transportation emulator (VTE) which emulates the whole packetized voice communications process including voice packetization, CODEC, speech detection, cell assembly, transmission error injection, packetized voice receiver, ATM cell disassembly and sidetone and echo emulation. The emulator is implemented on a DSP56000 ADM system; c) analog interface which provides an interface between analog telephone set and the speech emulator. It consists of AD/DA convertor, anti-alias filters and analog telephone interface circuit (audio interface). If necessary, the audio interface can also inject sidetones and generate echoes through its hybrid network.

The ATM network simulator and speech emulator are connected through an RS-232 interface. The cell delay and cell loss data generated from the network simulator is passed through the link to the speech emulator every 6 ms (packetization

Emulation Setup



Telephone:	modified model 100	Computers:	PC 386 Se
Audio I/F:	self designed		DSP56001 development system
L. P. F:	cutoff frequency 3,400 Hz		
AD/DA:	Ariel ADC56		

FIGURE 5.1 ATM Speech Transport Emulator Layout

period). The man-machine interface (MMI) is also implemented on the PC which provides the initialization and overall emulation control of the speech transport emulation. It is a fully menu-driven interface allowing an easy interaction with the system.

5.3 ATM Network Simulation

In ATM voice communication, packetization delay is either 6 ms (48 octets cell) or 4 ms (32 octets cell). The speech emulator has to be able to generate a cell, add on the cell delay and loss information to the cell and deliver it to the PVR in 6 or 4 ms. To meet such a requirement, the ATM network simulator (PC) needs to deliver the cell delay and loss information to the speech emulator every 6 or 4 ms based on its simulation results. Because of the processing power limitations, it is difficult to generate cell delay and loss information on-fly inside PC. That is why an alternative solution has to be taken. Cell delay and loss information are generated before emulation starts. The data is saved in two data files. One file contains the cell delay information of each individual cell which is generated by ATM network simulator. As discussed in Chapter 3, the cell delay process follows a Gamma distribution. The second data file contains Poisson distributed data of the cell loss intervals. During emulation process, both files are opened. The system initially reads a cell loss interval from the cell loss data file, then read the delay data file and send the delay data to speech emulator every 6 or 4 ms. In the mean time, the system also counts the number of cells having been transmitted. If the number of transmitted cell equals to the cell loss interval, the system will mark the cell as lost. Then the system reads another cell loss interval data and restarts to count the number of cell being transmitted. If the match occurs, the cell will be marked as lost cell again. This process continues until the end of the data file is reached.

As an emulation configuration option, the data files can be repeatedly used in one simulation. In other word, when the end of data file is reached the system does not stop the emulation, it rereads the data file from the beginning again in order to keep the emulation alive. In fact, if the repeat cycle (data file) is large enough, this will not affect the emulation result at all.

ATM network simulation contains the following major functions:

- User interface
- Emulation initialization and emulation control
- ATM network behavior simulation
- Cell delay and cell loss data file generation

5.3.1 User Interface

A full menu-driven user interface is implemented for full emulation control. It includes 4 sub-menus:

- Emulator Initial Configuration Menu
- Generate Simulation Data File Menu
- Simulation Data Selection Menu
- Emulation Control Menu

5.3.2 Emulation Configuration and Control

In emulation configuration process, following parameters can be initialized:

1). Cell size selection which is used to select either the North America proposed standard (53 octet) or the former European proposed standard (36 octet).

2). Delay compensation/reconstruction delay which is used to compensate cell delay fluctuations.

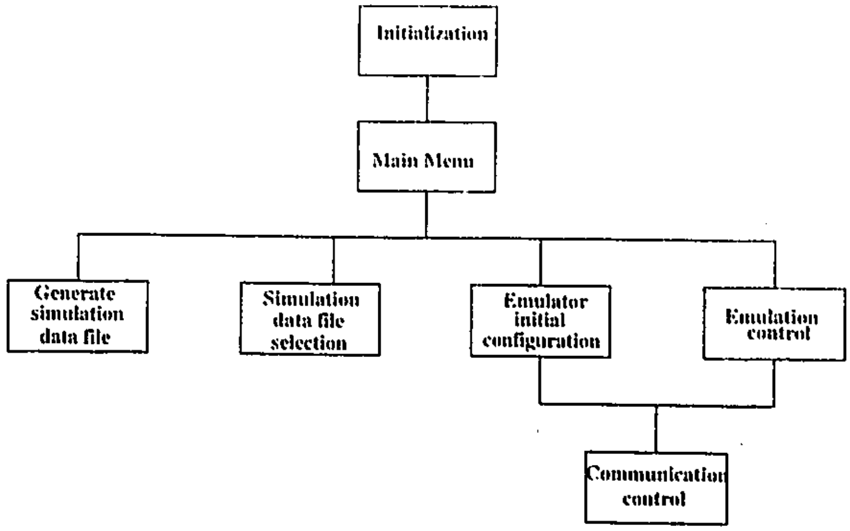


FIGURE 5.2 Software Hierarchy Structure of Main User Interface

3). Speech CODEC selection is used to select speech CODEC. In the current implementation only μ-law and A-law PCM CODEC have been implemented.

4). PVR algorithm selection. In the current implementation, three options have been implemented. Option 1: no delay compensation, PVR play-out cells as they arrive; Option 2: fixed delay compensation; Option 3: adaptive delay compensation.

Figure 5.3 gives the hierarchy structure diagram of the speech emulator configuration software. The input parameters are organized into several emulation control words, then sent to the DSP56000 ADM to get the speech emulator initialized.

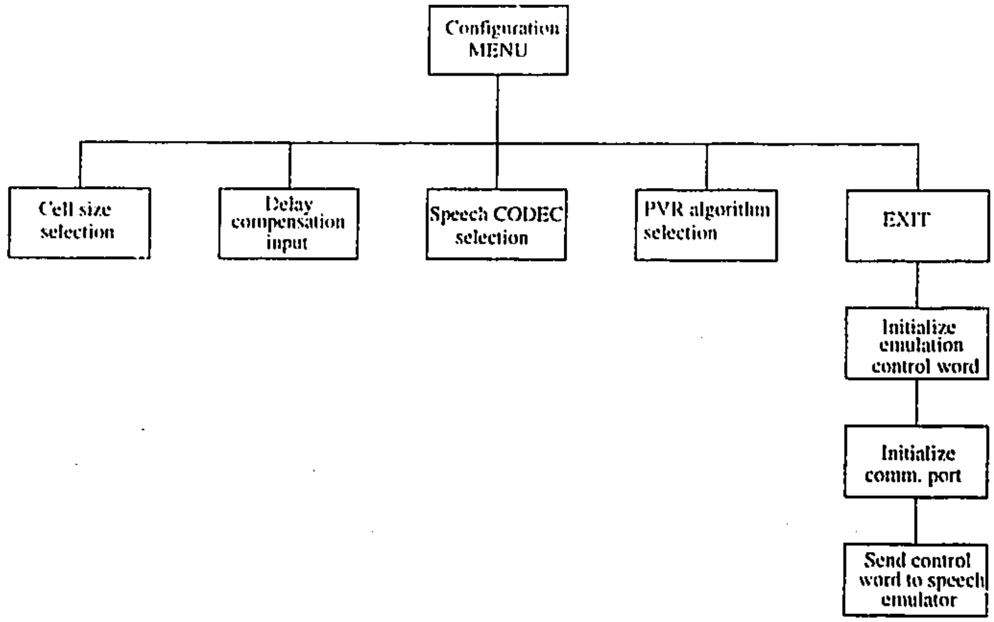


FIGURE 5.3 Software Hierarchy Structure of Speech Emulator Configuration

5.3.3 ATM Network Behavior Simulation

The major function of the ATM network simulation is to generate two data files which contain the random values describing the cell delay and cell loss properties. The cell delay data file is generated based on the distribution given in *EQ. 3 - 6*.

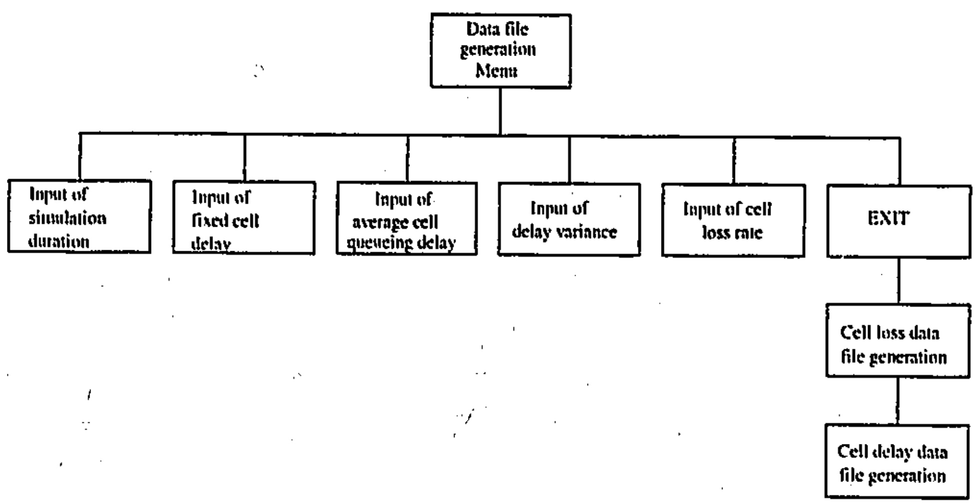


FIGURE 5.4 Software Hierarchy Structure of Data File Generation

Cell loss data file is generated according to the Poisson distribution which is given in *EQ. 3 - 7*

Figure 5.4 shows the structure of the random files generation process. In this module, the input parameters will be properly converted into the format that the random number generator can recognize. At the end of this process, two data file are generated and saved in the hard disk.

1). Simulation duration is used to calculate the size of data file. For a single pass simulation, it is the duration of the emulation. In this case, the data file is only used once. For the repeated emulation it is the duration of each emulation cycle.

2). Fixed cell delay is a summation of fixed delay a cell experienced in the network.

3). Average cell queueing delay is actually the mean of D_v (c.f. *EQ. 3 - 3* for the calculation of the mean).

4). Delay variance will decide the shape of cell delay *pdf*'s curve. (c.f. *EQ. 3 - 4*)

5). Cell loss rate is the ratio between cell loss and the total number of cell generated.

5.4 Design of Speech Emulation Process

The speech emulator has been built on using the Motorola DSP56000 ADM board with an Arial DSPAD16 card which has 2 16-bit A/D inputs and 2 16-bits D/A outputs used as analog telephone signal interface. The functionality of speech emulator can be summarized as follows:

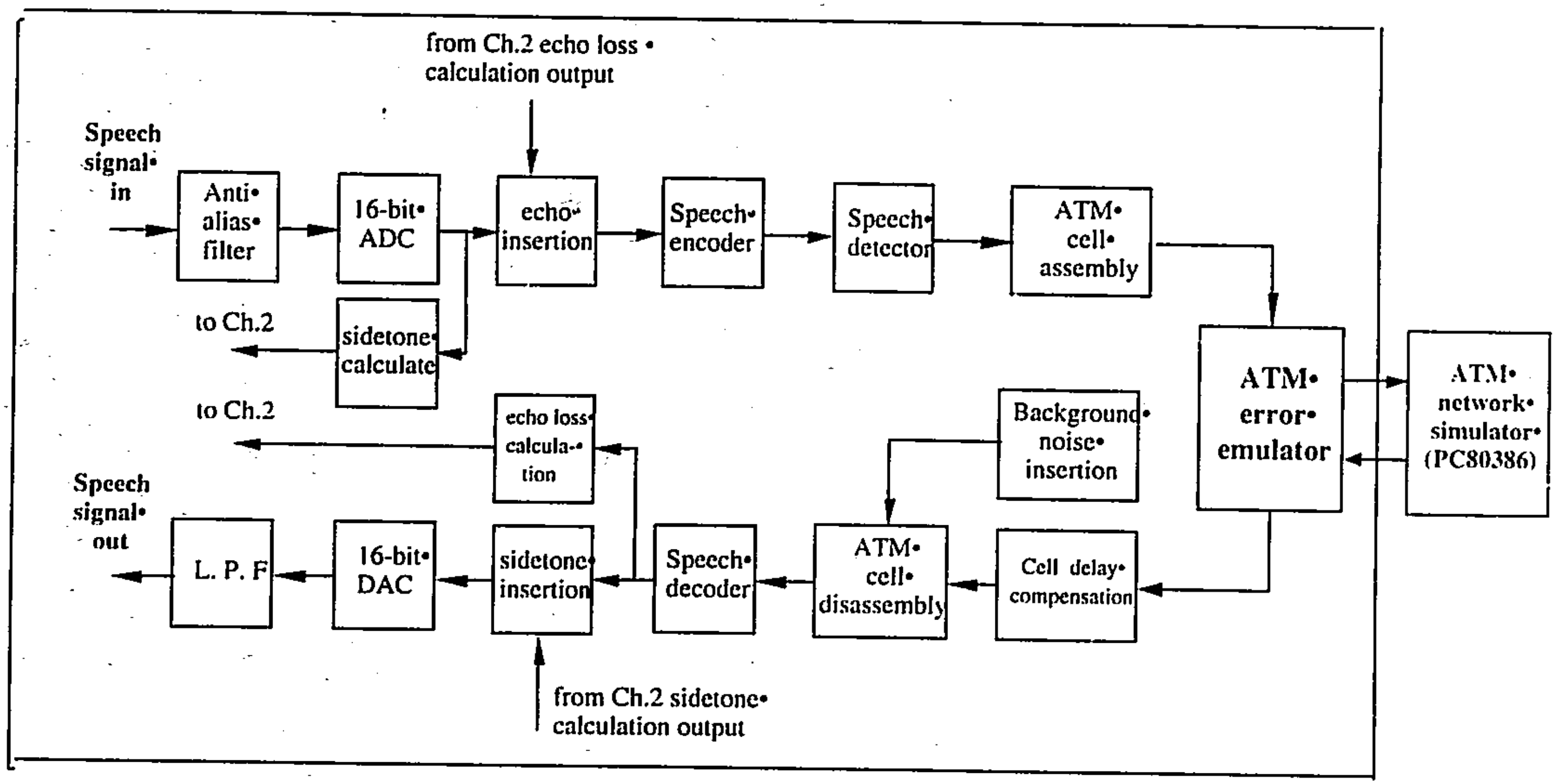


FIGURE 5.5 ATM Speech Transport Emulator Block Diagram

1). analog speech signal digitization

2). sidetone and echo insertion

3). speech/silence detection

4). ATM cell assembly(packetization) and PVT cell transmission emulation

5). packet voice receiver with cell delay compensation

6). cell disassembly

7). speech signal reconstruction

Figure 5.5 depicts the architecture of the speech emulator on one direction. The other direction is symmetrical. The high level software hierarchy structure diagram is given in *Figure 5.6 on page 71*.

5.4.1 Analog Speech Signal Digitization

The analog speech signal AD conversion is performed on the Arial DSPAD16 board. An 8 Khz timer is used as the trigger source to get the voice signal sampled at the speed of 8000 samples per second. Then the 16 bits voice samples are sent to µ-law or A-law speech CODEC on DSP56000 ADM card, which converts the voice samples into 64 kbits PCM bit stream. Because of modular software design, the CODEC can be easily replaced in future evolutions. When different bit rate CODECs are used the ATM packetization delay has to be changed accordingly. For example, if an ADPCM encoder is used, the packetization delay will be increased to 12 ms instead of 6 ms.

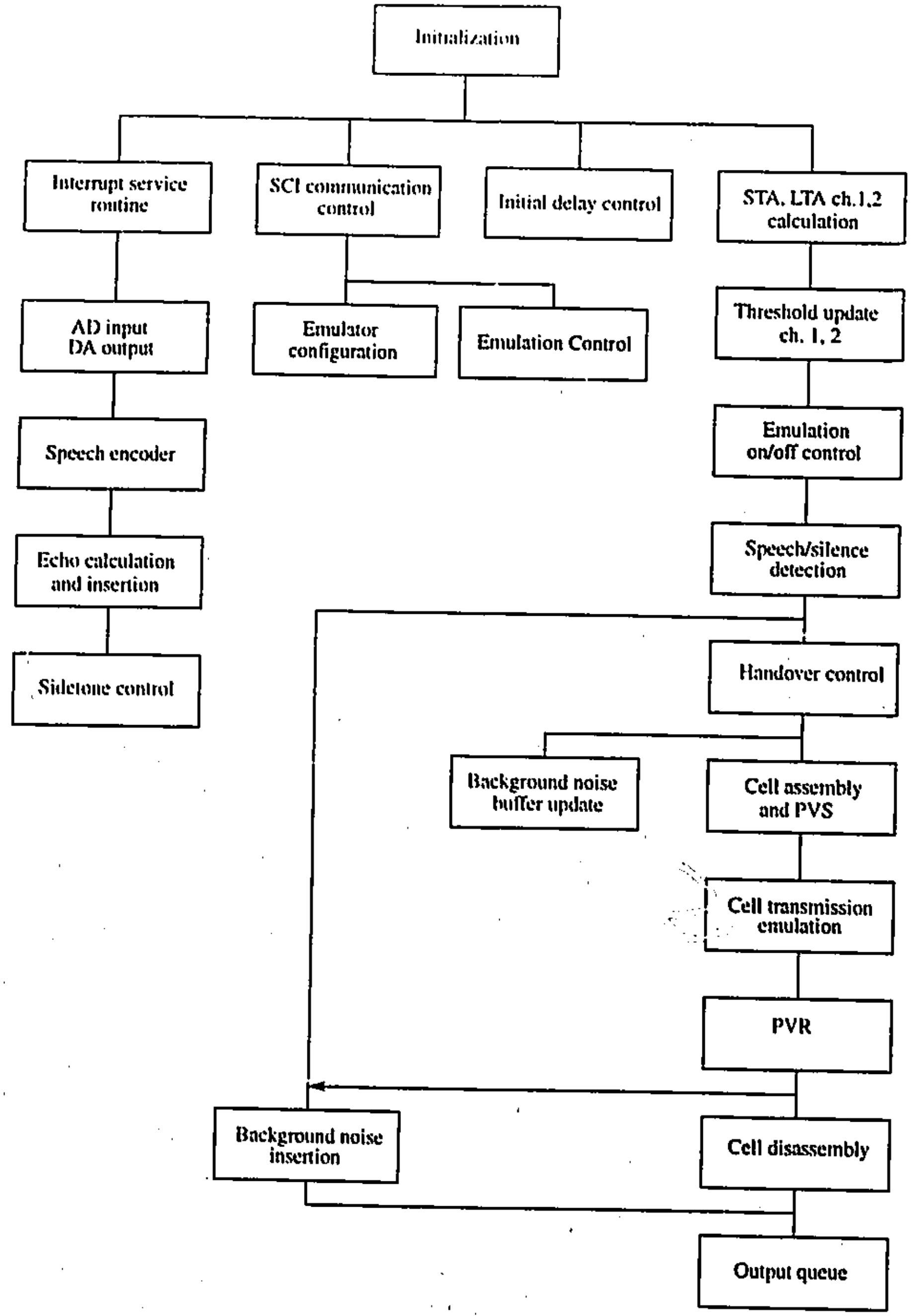


FIGURE 5.6 Software hierarchy structure of speech emulator

5.4.2 Sidetone and Echo Insertion

In a telephone set, partial of the transmitter signal power is diverted to the receiver in order to make the talker listen to his/her own voice while he/she is talking, which is defined as sidetone.

In general, there are two methods to add the sidetone and echo into the emulation system: hardware approach and software approach.

In the hardware approach, the analog hybrid circuit in the telephone set is used to generate sidetone and echo. Since the telephone set is specially designed to work with current loop, only a fixed level of sidetone and echo can be generated from the telephone hybrid circuit. This is definitely not good enough to meet the requirement of a speech emulation study.

At the earlier stage of the experiment, the circuit in a telephone set was used to provide sidetones and generate echoes during the emulation. It works fine for the fixed level sidetone and echo generation. As experiment progressing, it was noticed that the level of echo generated from phone set can only be adjusted in

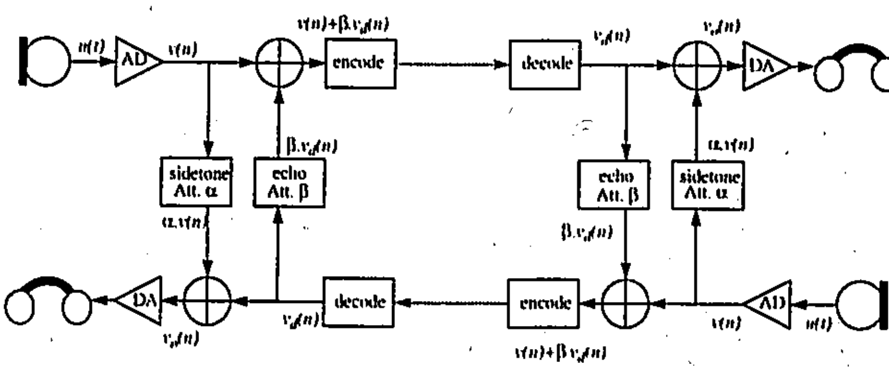


FIGURE 5.7 Sidetone and Echo Insertion on Audio Paths

about 3 dB range by changing the loop current. It is really inconvenient when trying to see the impact of different echo loss to the speech quality. Based on this motivation, a software approach has been designed and implemented on sidetone and echo insertion.

Figure 5.7 illustrates how the sidetone and echo are injected into the audio paths in the software approach. The sidetone signal is taken from the output of the AD convertor, then multiply with an attenuation coefficient α. The output signal $\alpha\ v_d(n)$ will be sent to the input of DA convertor combining with the output data $v_d(n)$. The echo signal is taken from the decoder output signal $v_d(n)$, then multiply with echo loss coefficient β. The attenuated signal $v_d(n)$ is combined with $v(n)$ as encoder's input signal: $v(n) + \beta \cdot v_d(n)$. Both α and β can be configured through emulation control interface either before or after the emulation is started.

5.4.3 Speech/Silence Detection

As already stated, the active talkspurt only takes about 40% of time in a conversation[53]. In ATM network, in order to increase the channel utilization, a speech/ silence detector is used to detect active speech signal.

After the speech signal is digitized, the 64kbits stream will go through the speech/ silence detector. Only active speech data is packetized into each 48 bytes ATM cell, then sent to the transmission emulation module.

In ATM speech emulator, the key technique of speech detection is the precise estimation of the background noise power during a pause or silence period. [41] By calculation of the silence noise power an adaptive power threshold for speech detection and a noise level code for silence reconstruction are obtained.

The signal power is estimated with a 8 ms moving window which has 64 speech samples of PCM coding, so it is possible to precisely detect the power variation of the speech waveforms. The short-time power P_n made by a low-pass operation method [44] using the previous power P_{n-1} and the input signal x_n can be calculated as:

$$P_n = \left(1 - \frac{1}{64}\right) \cdot P_{n-1} + \frac{1}{64} \cdot x_n^2 \qquad \text{(EQ 5 - 1)}$$

The speech/silence detection is made by comparing the power P_n with the adaptive power threshold P_{TH} which is given in *EQ. 5 - 3* , below.

The final decision of talkspurt or active speech duration t_A is the combination of the speech detection time t_V and the hangover time t_H as follows:

$$t_A = t_V + t_H \qquad \text{(EQ 5 - 2)}$$

The speech signal is delayed by 12 ms relative to the speech decision to minimize front-end clipping. This 12 ms delay is suitable to PCM speech packetization for an ATM cell size of 48 octets [2]. For the adaptation of P_{TH}, the silence noise power P_s is produced by two operational methods. Calculating the average power P_N during 256 ms, the minimum value of P_N is renewed as long as the silence duration is continued. This renewal P_N becomes P_s. If the speech duration is continued up to 5s, the minimum value of P_N within 5s becomes P_s. Using P_s with these operations, P_{TH} is calculated by:

$$P_{TH} = \alpha \cdot P_S \qquad \text{(EQ 5 - 3)}$$

where, the coefficient $\alpha = 10^{0.6}$ and P_{TH} range from -60 dBm0 to - 30 dBm0 [37].

To maintain speech quality, a longer hangover time has been proportionally applied to the power threshold increase. [42] Therefore control of the adaptive hangover time is carried out in the range of variable time from 20 ms to 160 ms.

A flow chart of speech/silence detection implementation of speech emulator is given in *Figure 5.8* .

Speech/Silence Detection

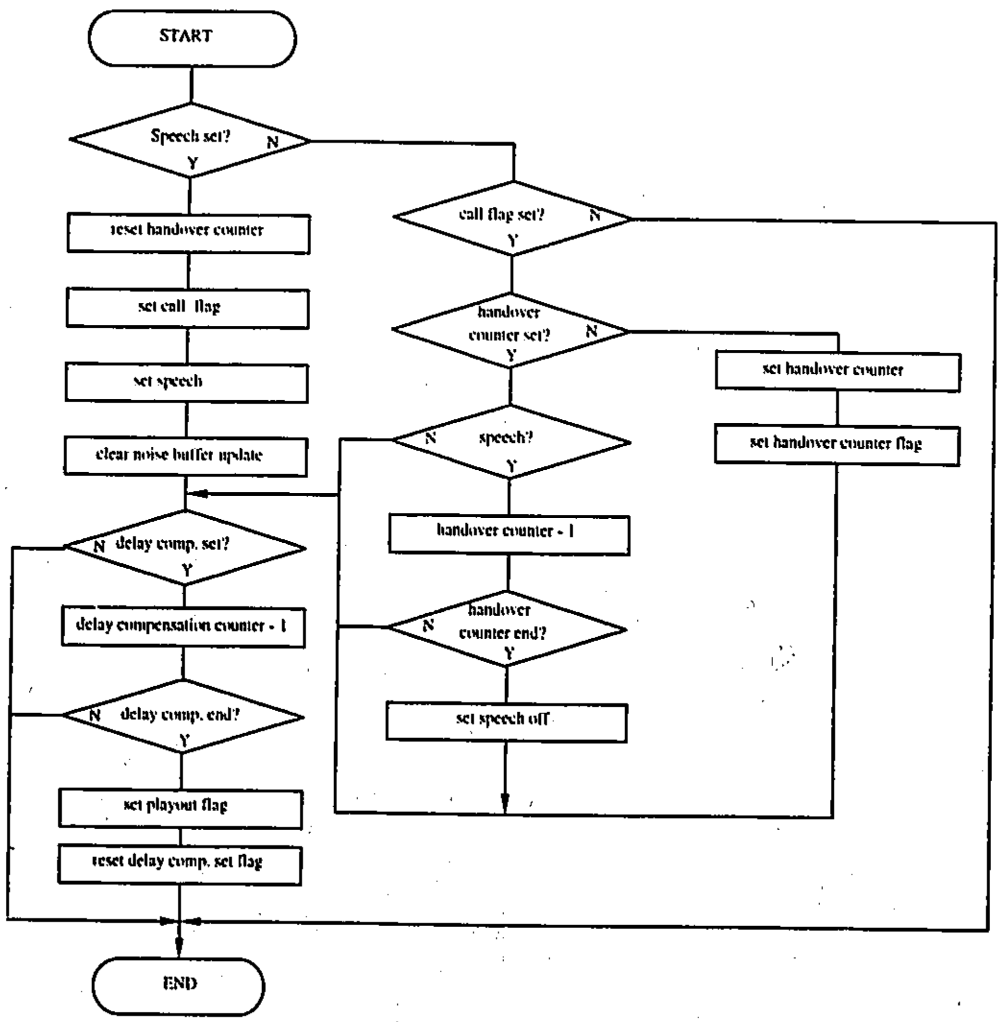


FIGURE 5.8 Flowchart of Speech/Silence Detection

5.4.4 Cell Transport Emulation

The active speech samples out of the speech detector are directly packetized into either 32 octets or 48 octets cells [30], which is based on the emulation configuration. In current implementation, one ATM cell only contains the voice data from one source. To fill up a 48 octets cell, it will take 48 speech samples (6 ms).

As soon as a cell is assembled, the cell will be sent to the transmission error simulator (TES). TES receives the network simulation data every 6 ms which is designated for each individual cell. If the data indicates (0xFFFF) the cell is a lost cell, then TES discards the cell. Otherwise, it attaches the data received from network simulator, then passes the cell to packet voice receiver. The data indicates the delay of the cell experienced in the network.

5.4.5 PVR Implementation

Packet voice receiver is the core of the emulator. It receives the cells from the network, estimates the delay cell experienced in the network, calculates the reconstruction delay and disassembles the cells, then puts the speech data into the playout buffer queue. If a cell is lost or experiences an excessive delay in the network, PVR inserts the background noise to the output buffer queue in order to make the output speech signal continued and sound more real.

In the current design, two types of packet voice receivers are implemented. The first one is a fixed delay compensation (FDC) PVR (please refer to *Figure 4.5 "Fixed delay compensation ATM PVR"on page 56* for detail), which uses the Blind Delay method (*Section 4.3 "Packet Voice Synchronization"*) to estimate the maximum queueing delay in the network and uses NTI strategy (*Section 4.4.1 "Null Timing Information (NTI) Strategy" on page 53*) to synchronize the speech. The second one is an adaptive delay compensation (ADC) PVR (Please refer to

Figure 4.6 "Adaptive Delay Compensation PVR" on page 58), which adjusts the compensation delay, between two consecutive talkspurts, according to the cell arrival variance in the previous talkspurt. The adjustments (t_{c1}, t_{c2} in *Figure 4.6 on page 58*) could be either positive or negative. If it is a positive adjustment, it means that the previous delay compensation is not long enough and the extra delay t_{ci} shall be added to the previous compensation delay *T*. The compensation delay after adjustment is given by $T_{new} = T + t_{ci}$. In this case, the proper number of bytes of background noise data are inserted into the output buffer queue before the actual speech data. If the adjustment is a negative, it means that the previous compensation delay is too long and it should be reduced according to the value of t_{ci}. In this case, PVR reduces the silence period between two talkspurts by shifting the speech data in the play-out buffer forward according to the new compensation delay $T_{new} = T_{pre} - t_{ci}$. For both PVRs, the initial compensation delay *T* is initially introduced by the user together with other configuration parameters before starting the emulation process. During the emulation process, these numbers can also be changed based on the experiment requirement.

Because of real time limitations, we could not calculate the cell interarrival variance in real time. Instead, we use the maximum cell play-out delay difference *MAXPD* multiplied by a safety coefficient α ($\alpha > 1$) to calculate the delay adjustment. *EQ. 5 - 4* determines the rules to calculate the delay compensation adjustment. It covers three scenarios. In the first scenario, *MAXPD* is larger than D_p, which indicates that one or more cells may have been lost or excessively delayed in the network. In this case, ($MAXPD - nD_p$) is used for the calculation of delay compensation adjustment. N is the upper bound on the number of consecutive cells can be lost in a talkspurt without being taken as a silence period. Based on the values given in *Table 2.1* , $N = 40$ which is defined as the maximum cell arrival delay in a talkspurt has been recommended in this study. It is equivalent to the mean of 240 ms silence period of a conversation. If a cell arrival delay is larger than 240 ms, it is considered as the first cell of a following talkspurt. In the second scenario, *MAXPD* is larger than 0 and smaller than the cell packetization delay D_p. In this case the delay compensation adjustment $t_{ci} = \Delta T = \alpha \cdot MAXPD$ is a positive number, and T will be extended by ΔT period. In the third scenario, *MAXPD* is smaller than 0 and $\Delta T = MAXPD + \frac{|MAXPD|}{\alpha}$ is a negative number, and T will be shortened by a period length ΔT. For example, if *MAXPD* equals to 3 ms, it indicates that in the preceding talkspurt, at least, one cell has arrived exceeding its target play-out time by 3 ms. In order to avoid excessive cell losses caused by extra cell delay in current talkspurt, it is recommended to multiply *MAXPD* by a safety coefficient α in order to get the actual delay compensation adjustment. *EQ. 5 - 4* gives the rule that the compensation delay T is adapted based on α and maximum play-out delay difference (*MAXPD*) measured in previous talkspurt and coefficient α. In fact, α is a parameter which may vary according to the change of cell delay variance and cell discarding rate. Because the limitation of CPU's processing power, only a fixed α = 1.6 was considered in the implementation.

$$T_{new} = \begin{cases} T + \alpha \cdot (MAXPD - nD_p), & MAXPD \geq nD_p, \quad n = 1, 2, \ldots M \\ T + \alpha \cdot MAXPD, & D_p > MAXPD \geq 0 \\ T + \left(MAXPD + \frac{|MAXPD|}{\alpha} \right), & MAXPD < 0 \end{cases} \qquad \text{(EQ 5 - 4)}$$

The algorithm of the adaptive PVR is given in *Figure 5.9* in the form of flow chart. In *Figure 5.9* , 1CEL_SPURT represents the first cell in a talkspurt. If the cell is the first cell in a talkspurt, PVR will update its compensation delay T according to the maximum play-out delay difference (MAXPD). P_Delay is the target play-out delay for current talkspurt. It is given by adding the delay experienced by the first cell in the talkspurt to T_{new}. This target play-out delay is used as the reference to judge the loss of other cells in the talkspurt. The first cell in a talkspurt is always played-out with an extra T ms delay. After the new delay compensation parameter T_{new} has been derived, the pointer to the play-out buffer is adjusted based on the difference of T_{new} and T, which is defined as ΔT. In the figure, ΔT is converted to the number of bytes N which is used to off-set the pointer of the play-out buffer. If T_{new} is different from the previous T, N bytes noise data will be either inserted into play-out buffer at the case of $\Delta T > 0$ or deleted from the play-out buffer at the case

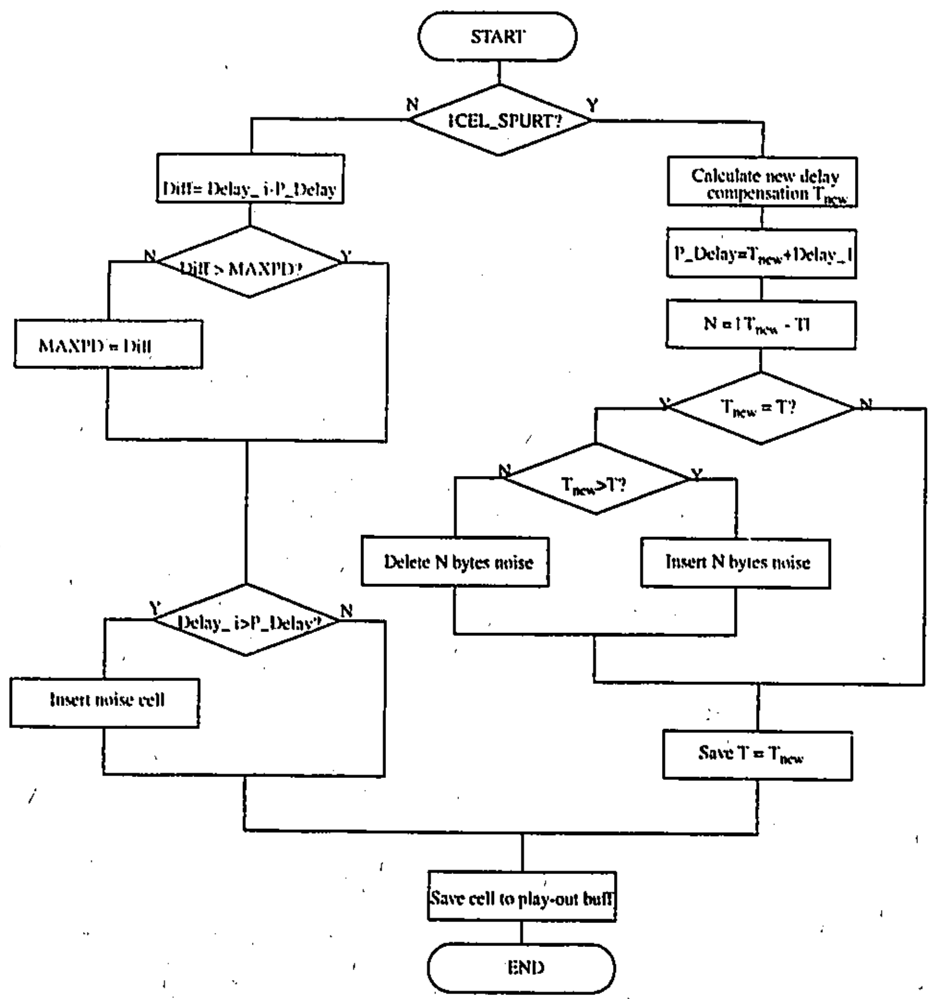


FIGURE 5.9 Adaptive Packet Voice Receiver Algorithm

of $\Delta T < 0$. Otherwise, no adjustment is necessary. If the arriving cell is not the first cell in the talkspurt, the PVR calculates the difference (*Diff*) between the cell delay (*Delay_i*) and the cell play-out target delay (*P_Delay*). If this difference, Diff, *Diff* is greater than the maximum value already registered, *MAXPD* is updated. If *Diff* is positive, the cell is dropped and one or more noise cells are inserted into play-out buffer according to cell arrival delay. Otherwise, the cell is saved in the play-out buffer until decoded.

5.4.6 Background Noise Insertion

Since the speech detector is used in ATM network to increase the channel utilization, there will be nothing transmitted from the source between two active talkspurts on the network. This silence period compression causes differences between transmitted and reconstructed periods at the receiving end when the silent periods are reconstructed[5]. In order to make the speech reconstruction more natural, the background noise should be added in the silence period.

In the real system, the PVR should generate the background noise according to noise level information provided by the transmitter end. In the emulator implementation, in order to reduce the design complexity, the background noise during silence period at the transmitter end is saved into a background noise buffer which will be keeping updated. As soon as PVR detects a cell loss or a silence period, it copies the noise data from this buffer into the voice data output buffer to fill out those blank periods.

5.4.7 SCI Communication on DSP56000

The SCI port on DSP56000 has been used to provide the communications between ATM Speech Transport Emulator (DSP56000, ADM) and ATM Network Behavior

Simulator (PC). The parameters passed through the link are emulation configuration, emulation control and network behavior parameters.

The SCI port has been set at the bit rate of 19.2 kbits per second. Because the ADM card does not have enough timers to support both 8 Khz speech signal sampling and 19.2 kbps serial communications, an external clock has to be used to provide the timing signal for the SCI. The RS-232 communication card on PC has been modified to output the clock signal on pin 16 as specified in CCITT V.24.

SCI port ADM board can only provide 0-5V signal level. In order to make the communications possible between DSP56000 and PC, a signal level conversion circuit has designed and peggy-backed on ADM board to convert the signal level between SCI and RS-232. The schematic of the converter is given in *Figure 5.10* for reference. The jumper should be removed if the internal timing source is selected.

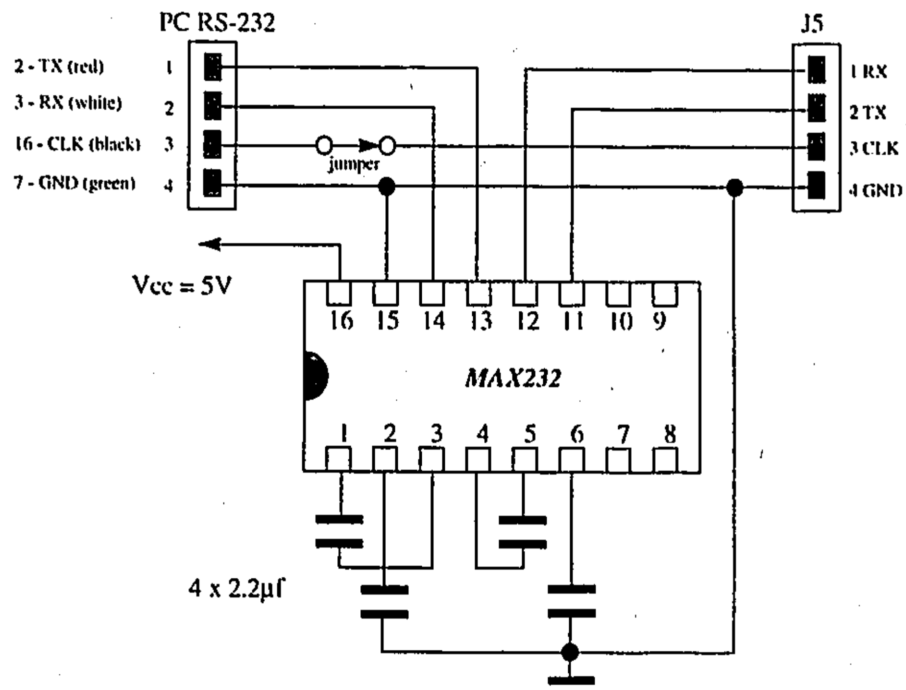


FIGURE 5.10 SCI - RS-232 Signal Level Convertor

5.5 Design of the Analog interface

The analog interface includes three subcomponents: a) human-machine interface which converts the acoustic speech signal into electrical signal. Currently two modified 100 type traditional telephone sets are used for the subjective test; b) an anti-alias filter which is used to reject any out of band signal; c) audio interface providing the interface between the telephone set and the AD/DA convertors.

5.5.1 Human Machine Interface

One of the objectives to build the speech emulator is to get subjective evaluations of the speech quality in ATM network. To achieve this objective, a human-machine interface is needed to convert acoustic voice signal into electrical speech signal at the talker end and convert the electrical speech signal back to acoustic voice signal at the listener end. Two 100 type regular telephone sets have been chosen for the purpose of the experiment.

Because the regular telephone set is designed to work on current loop with switch line card, it can not be directly interface to the AD/DA convertor. An audio interface needs to be designed and implemented to provide such kind of interconnection. On the other hand, the telephone set is designed to work in 2-wire full duplex mode, but AD/DA channel are equivalent to 4-wire duplex mode. To avoid designing a complex 2-wire to 4-wire convertor, the wiring in telephone set has been modified and to make it workable with 4-wire system. In the mean time, the hybrid circuit in the telephone set is kept the same, which still provide the sidetone to the receiver and also generate echoes to the emulator. The modified telephone circuit is illustrated in *Figure 5.11* .

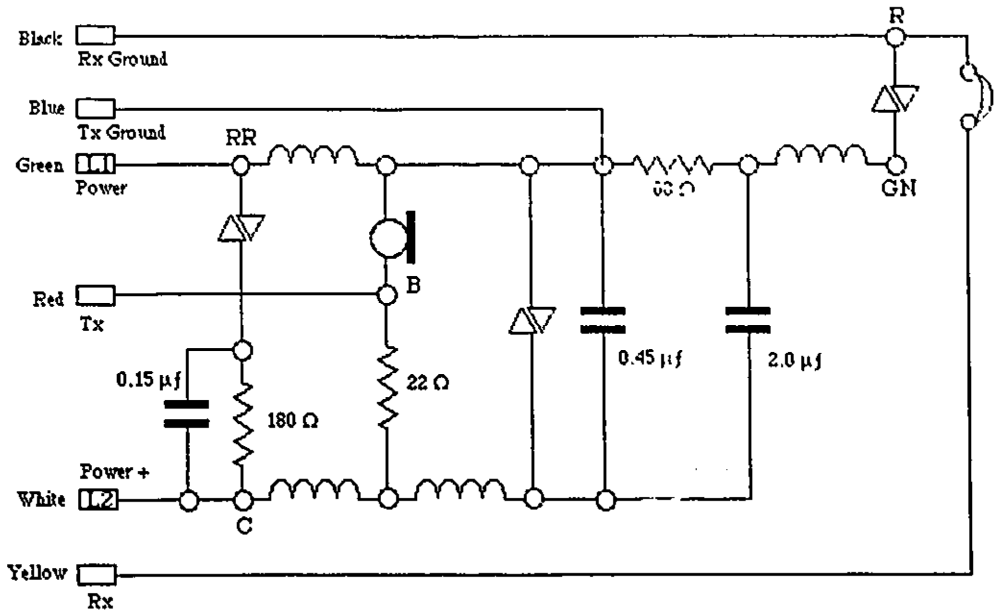


FIGURE 5.11 Modified 100 Type Telephone Set Circuit

5.5.2 Audio Interface

The main purpose of designing the audio interface is to provide a signal connection between the telephone set and AD/DA convertor. It has following basic functions:

- provide the current loop and power to the telephone set
- provide a way of flexible gain control between the phone set and the AD/DA convertors in order to meet the needs of experiment
- impedance conversion between DA and the phone set

5.5.3 Anti-alias Filter

Anti-alias filter is used to cut-off out band signal in order to reduce the interference to the analog-to-digital convertor. A 7th order Elliptic low pass filter, with 3 dB cutoff frequency at 3.5 Khz, has been selected and implemented on both AD inputs and DA outputs to protect the AD/DA circuit. The schematic of the low pass filter is given in *Figure 5.12* for reference.

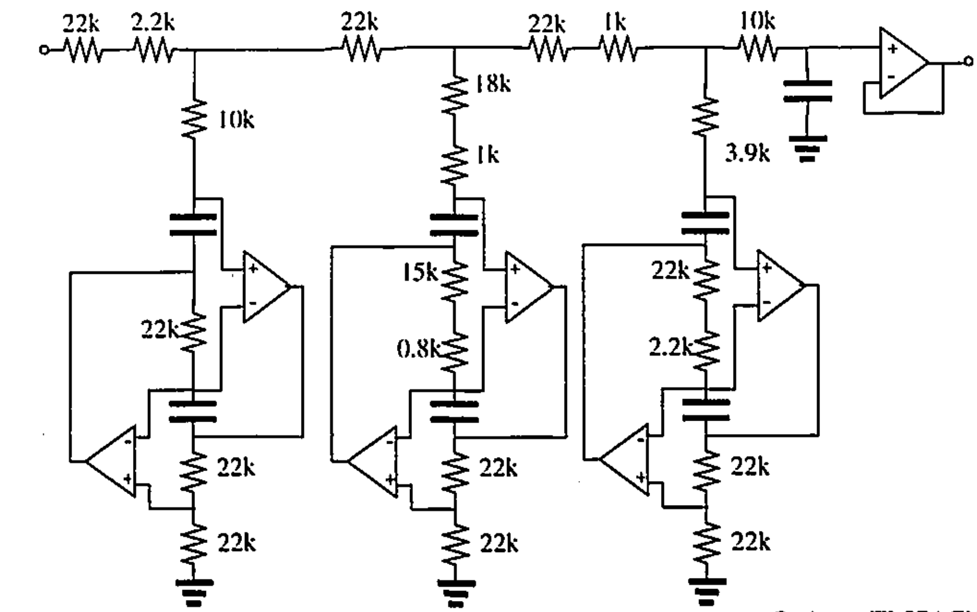


FIGURE 5.12 Schematic of Anti-alas Filter

5.5.4 Calibration

The analog portion of the system has to be calibrated before the emulation starts, in order to achieve the maximum signal to noise ratio at the AD input.

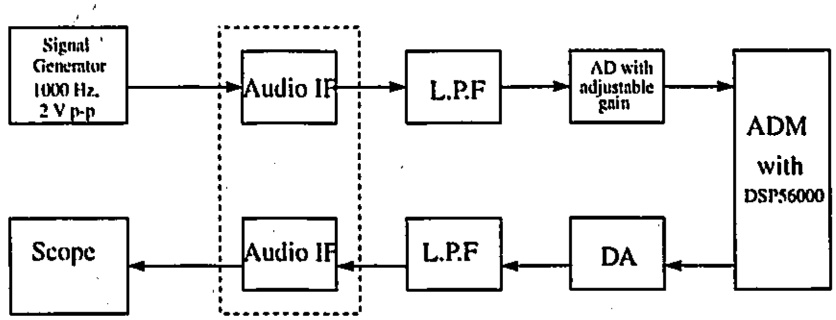


FIGURE 5.13 Calibration Block Diagram

In the calibration procedures, both a signal generator and a scope are used to find out the best gain which can achieve the maximum output with no distortion at the peak input.

The peak input signal can be calculated based on following method. Let's assume the average speaker power is -11.1 dB0 +/- 4.7dB[3] (1 mv = 0 dB0). This means that peak speaker's power will be -6.4 dB. According to the *EQ. 5 - 5*

$$P_{avg} = 10 \cdot lg\left(\frac{u^2}{r}\right) \qquad \text{(EQ 5 - 5)}$$

where: P_{avg}: average speaker power

u: average speaker's *RMS* voltage

r: input circuit impedance

then, the *RMS* speaker's signal voltage will be given by following calculation.

$$u = \sqrt{\frac{6}{10} \cdot 10^{-0.64}} = 0.37 = 370\,(\mathrm{mv})$$

Then the peak voltage will equal to $\sqrt{2} \cdot u = 523\,(mv)$. For safety reasons, 1000 Hz, 1 volt or 2 volt peak-to-peak signal is taken as the emulator calibration signal.

5.6 Summary

This chapter presented an overview of the design and implementation of the ATM speech emulator. Because the speech detection and background noise insertion techniques have been considered in the design of the speech emulator, it makes the operation of the emulator more close to reality. In next chapter, the speech emulator will be tested and evaluated through the ATM speech communication emulation studies.

CHAPTER 6

Experiments and Results Analysis

As discussed in earlier chapters, the main objective of building a speech emulator is to study of the speech signal transmission over ATM networks. In this chapter we use the ATM Speech Emulator designed in Chapter 5 to perform some experiments on ATM speech communications. Through these experiments, the functionality of the speech emulator is verified and some issues on speech signal transmission in ATM network are discussed and investigated as well.

A subjective speech quality evaluation method, the MOS score system as described in *Section 2.1.5 "Speech Quality Evaluation" on page 19* is adapted to our experiment. Three people have taken part in the subjective evaluation of speech communications under different conditions.

Subjective assessment results of speech quality were expressed in terms of an opinion equivalent Q (Qop) and MOS scores. The MOS score was first obtained as the average ratings the subjects gave to the speech quality, and then converted into opinion equivalent Q according to the curves given in *Figure 6.1* [45]. The opinion equivalent Q was defined as the speech to speech-correlated noise ratio of a reference signal.

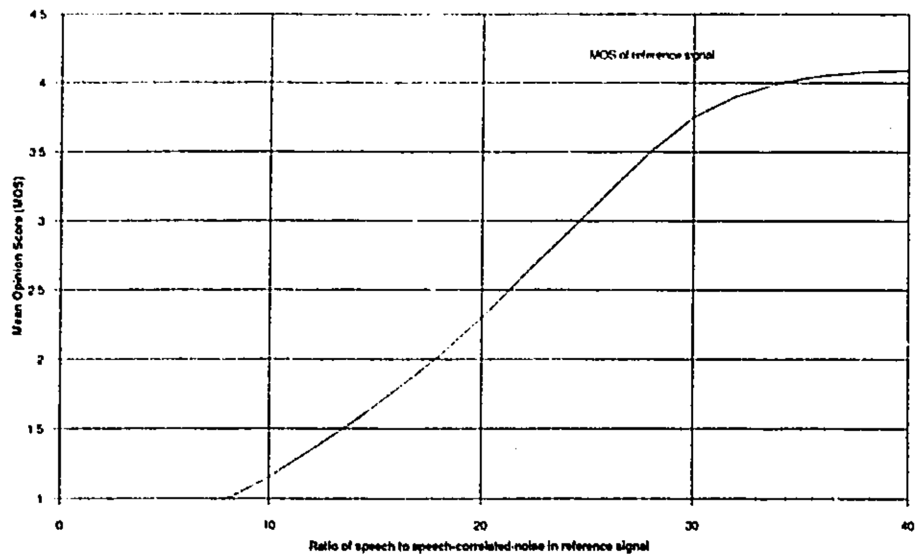


FIGURE 6.1 Determination of Opinion Equivalent Q (Qop), dB

A total three experiments were planned. The first experiment aimed at the study of speech quality degradation caused by cell loss rate and delay. The second experiment attempts to study the relationship between echo path loss and the impact speech transmission quality. The third experiment has an objective to compare the performance of the different packet voice receiving algorithms proposed.

6.1 Speech Quality at Different Cell Loss Rate

Three experiments were carried under in this category. During these tests, the echo was not inserted. The objective of the first experiment was to investigate the impact to the speech quality by missing cell. The second experiment was to investigate the speech quality degradation caused by different reconstruction delays. In the third experiment, we investigated the speech quality degradation caused by both missing cells and delay variation.

6.1.1 Speech Quality Degraded by Cell Loss

6.1.1.1 Experiment Conditions

The experiment was carried out in real time on the emulator under the following conditions

- Cell delay caused by network traffic conditions is Gamma distributed with: Mean = 10 ms and variance = 10 (*c.f. Figure 6.2*).
- Fixed network propagation delay is 50 ms.
- Reconstruction delay is set to 30 ms which is enough to compensate nearly 100% of late arrival cells. It means that missing cell caused by network queueing delay is close to zero.
- Cell size is either 32 octets or 48 octets.
- μ-law is selected as CODEC
- PVR type 2 - fixed reconstruction delay PVR is selected.
- Sidetone is inserted at 12 dB loss and no echo is inserted
- Emulation duration was 5 minutes for each evaluator at each cell loss rate level for each cell size.
- Each individual test was repeated three times for each evaluator.

During the experiment, all the parameters mentioned above were kept constant, which the cell loss rate was changed from 0 to 10%. The speech quality was not tested when the cell loss rate exceeding 10% because of poor speech quality.

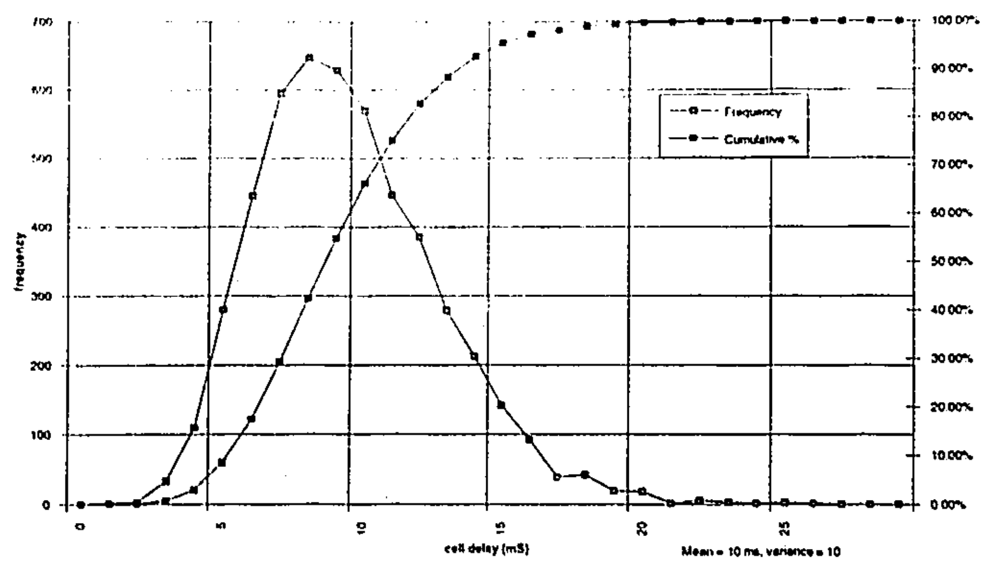


FIGURE 6.2 Simulated cell delay distribution (Mean = 10 ms, Variance =10)

6.1.1.2 Experimental Results

The experiment results are given in *Table 6.1*. It lists both MOS and Qop scores. The original scores were obtained in MOS form and converted into Qop form by using *Figure 6.1* . *Figure 6.3* and *Figure 6.4* are used to illustrate the difference on the results obtained for the 32 octets cell size and 48 octets cell size.

TABLE 6.1. Quality degradation by cell loss rate

	32 octets cell		48 octets cell	
Cell Loss%	MOS	Qop (dB)	MOS	Qop(dB)
0	4.06	36.78	4.05	36.57
0.1	4.02	34	4	33.8
0.2	3.95	32.67	3.92	32.26
0.5	3.82	30.67	3.75	30
1	3.48	27.7	3.4	27.3
2	3.02	25	2.95	24.3
5	2.13	18.46	2.01	17.86
10	1.65	15	1.45	13.07

6.1.1.3 Experimental Results Analysis

From *Table 6.1* and *Figure 6.3* , we see that the service quality is degraded as the cell missing rate increases. When cell missing rate approaches 3% in both 32 octets and 48 octets cell cases, the voice quality is obviously getting worse and worse. When cell loss rate is over 3%, the voice quality drops dramatically to the unacceptable level (MOS < 2).

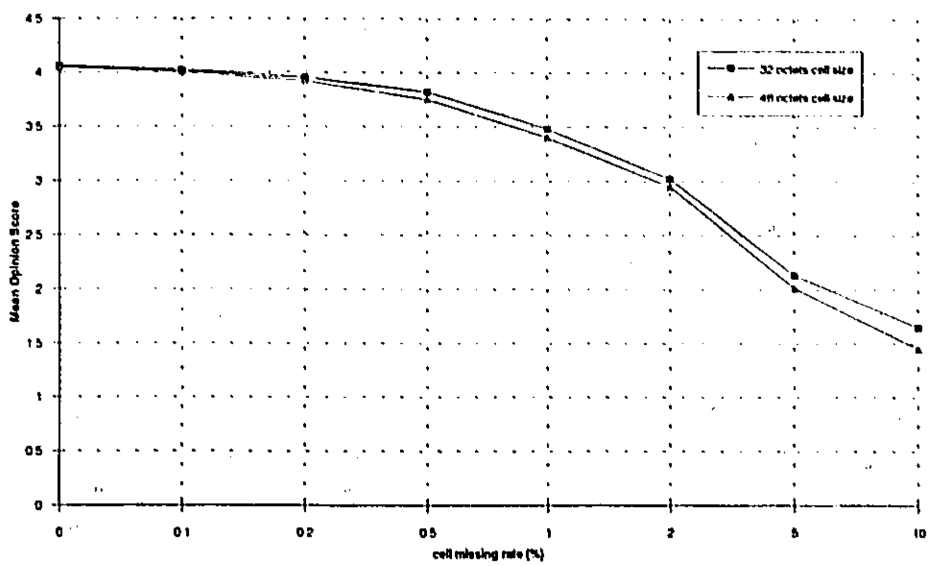


FIGURE 6.3 Mean Opinion Scores on cell missing rate test

There is a small difference on the voice quality between 32 octets cell and 48 octets cell at the same cell loss rate. During the experiment, we have noticed that the volume of clicks in 48 octets cell test is a little louder than the volume in 32 octets cell. This is obviously due to the fact that more information is lost. The same explanation is used to justify the articulation difference between these two tests as indicated in *Figure 6.4* .

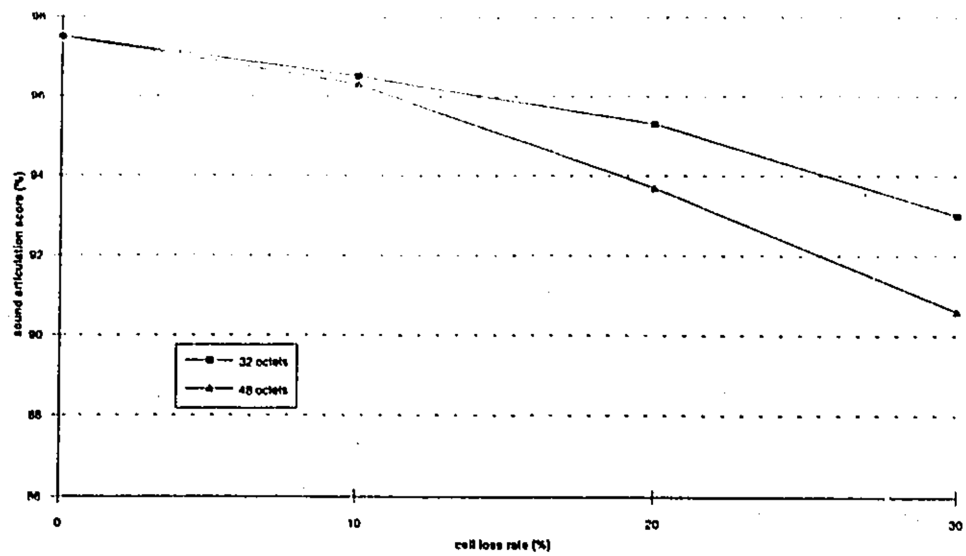


FIGURE 6.4 Sound articulation score on cell missing rate test

6.1.2 Speech Quality Degraded by Delay Compensation

6.1.2.1 Experimental Conditions

In this experiment, the tests were run using different reconstruction delays in order to investigate the speech quality degraded by the missing cell caused by cell delay. All the experimental conditions defined in *Section 6.1.1.1* are kept the same except for the following two conditions:

- reconstruction delay is varied from 10 ms to 25 ms in 5 ms granularity
- cell loss rate was set to zero

6.1.2.2 Experimental Results

The experiment results for different reconstruction delay are given in *Table 6.2* and *Figure 6.5* . The cell loss percentage labeled on the figure is the measurement result of actual cell loss rate under the given reconstruction delay. When reconstruction delay equals to 10 ms, the speech quality is very bad. That has been indicated by the MOS < 1 only.

TABLE 6.2. Quality degradation by different reconstruction delay

Reconstruction Delay (ms)	32 octets cell		48 octets cell	
	MOS	Qop (dB)	MOS	Qop(dB)
10	< 1	0	< 1	0
15	2.1	18.8	2.05	18.1.7
20	3.82	32.67	3.75	32.26
25	4.00	36.67	3.9	35.5

6.1.2.3 Experimental Results Analysis

In this experiment, because different reconstruction delays have been chosen, the cells whose queueing delay exceed the compensation capability of the reconstruction delay are lost. In this case the cell loss is purely caused by the variation of cell arrival delay. The equivalent cell missing rate at different reconstruction delay can be obtained on *Figure 6.2 "Simulated cell delay distribution (Mean = 10 ms, Variance =10)"on page 90*. For example, if the reconstruction delay is 18 ms (*c.f. Table A.1*), then the equivalent cell loss rate will be 1%. It has been noticed that the voice quality degradation caused by cell delay variation is worse than the degradation caused by random lost cells.

In ATM network, all cells from one voice source will go through one virtual path. As discussed in *CHAPTER 3*, the delay of a cell in a talkspurt can not have a delay shorter than its preceding cell by the delay of its preceding cell minus cell packetization delay. If one cell experiences a serious delay in the network, it may cause several succeeding cells to be lost at the packet voice receiver. For the same average cell loss rate, it will cause more bursty cell losses in this experiment than in previous experiment. That is why the MOS score is lower in this case. According to *Figure 6.5* , when reconstruction delay is 18 ms and equivalent cell loss rate equals to 1%, the MOS score is about 3.1 instead of 3.4 in previous test.

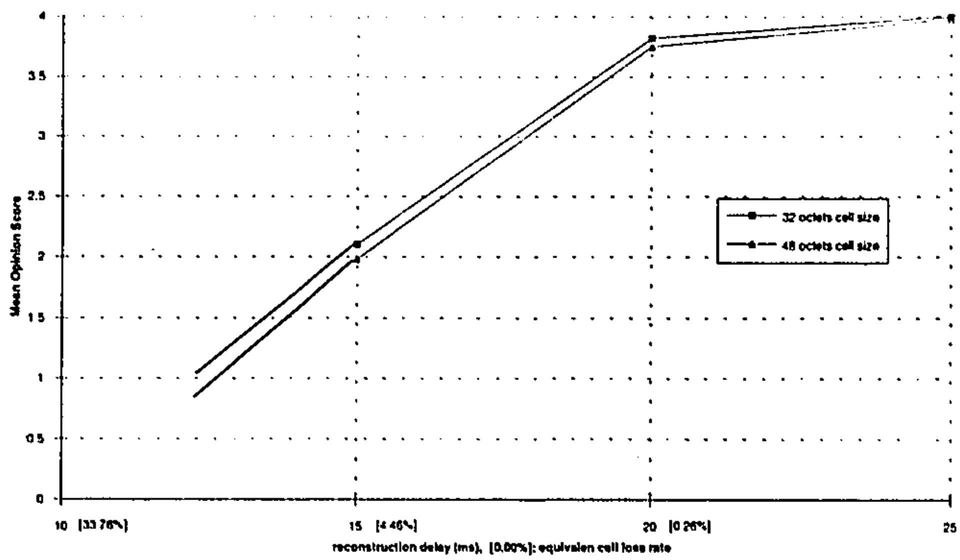


FIGURE 6.5 Mean Opinion Score on different reconstruction delay

6.1.3 Speech Quality Degraded by both Cell Loss and Cell Delay

6.1.3.1 Experimental Configuration

In this experiment, the objective was to find out how the cell missing rate and cell delay variation degrade the speech quality altogether. The test was run under a fixed reconstruction delay which introduced about 1% cell missing rate, and a variable cell loss rate from 0 to 10%. Most of the experiment conditions are the same as defined in *Section 6.1.1.1* except the reconstruction delay is reduced from 30 ms to 18 ms. It introduces 1% cell loss to the total cell loss rate, which is caused by excessive cell delay.

6.1.3.2 Experimental Results

Table 6.3 gives the experimental results of the speech quality degradation caused by the cell loss from both random discarded cell and extensive cell delay. The cell loss rate on the left column in *Table 6.3* counts only random discarding cell loss rate, which does not include the cell loss caused by cell delay. The cell loss rate caused by exceeded cell delay is about 1% at 18 ms reconstruction delay (*c.f. Table A.1*).

TABLE 6.3. Quality degradation by both cell loss rate and cell delay

	32 octets cell		48 octets cell	
Cell Loss%	MOS	Qop (dB)	MOS	Qop(dB)
0	3.48	27.7	3.4	27.3
0.1	3.4	27.2	2.95	24.3
0.2	3.3	26.7	2.01	17.86
0.5	3.2	26	1.45	13.07
1	3.01	25	2.95	24.3
2	2.64	22.1	2.01	17.86
5	1.9	17	1.45	13.07
10	1.65	15	1.45	13.07

6.1.3.3 Experimental Results Analysis

In this experiment, both random discarding cells and exceeded cell delay have been considered in the test and the MOS score is better than experiment 2, but worse than experiment 1. Better bursty cell losses comparing with experiment 2 will explain the reasons.

According to these experiments results, it can be concluded that the bursty cell losses caused by cell delay variation will degrade the speech quality more than random discarded cell loss. In real ATM system, the cell delay variation caused by network queueing delay is not stable. It varies dramatically as network traffic condition changes. This could be reflected in our emulation by changing the cell delay

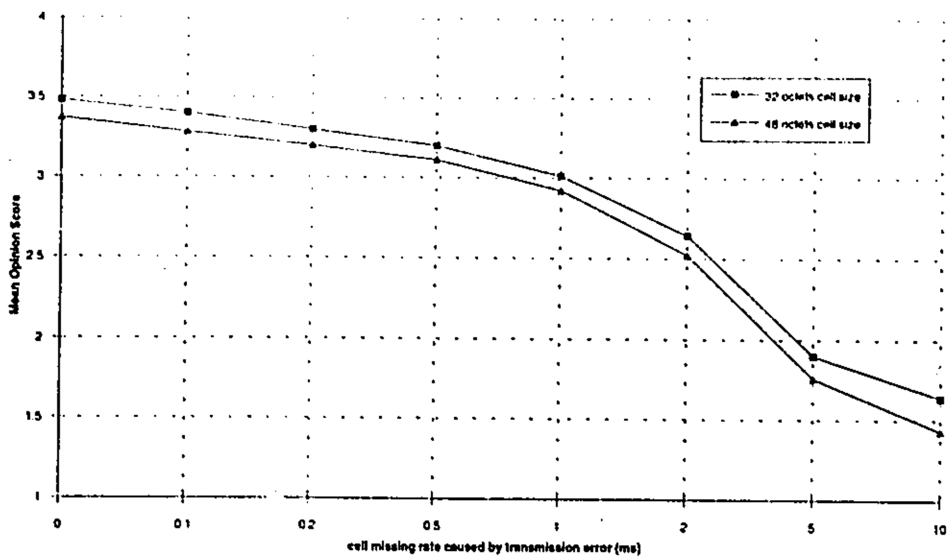


FIGURE 6.6 Mean Opinion Score on cell loss rate at reconstruction delay = 18 ms

variance. *Figure 6.7* displays the relationship between cell delay variance and PVR reconstruction delay. It is obvious from the figure, when cell delay variance increases, the reconstruction delay has to be increased proportionally. Otherwise, more cells will be discarded at the PVR because the delay compensation is not long enough. This reminds us that the reconstruction delay should be selected based on the maximum cell delay variance in the network in order to avoid unnecessary cell loss. This is not a efficient approach. But it is safe. Otherwise, an adaptive delay compensation algorithm has to be considered in PVR design.

The experiment results obtained from test 1 are a close agreement to the results obtained by N. Kitawaki [45] and H. Nagabuchi [46] which proved that the functionality of the speech emulator is correct.

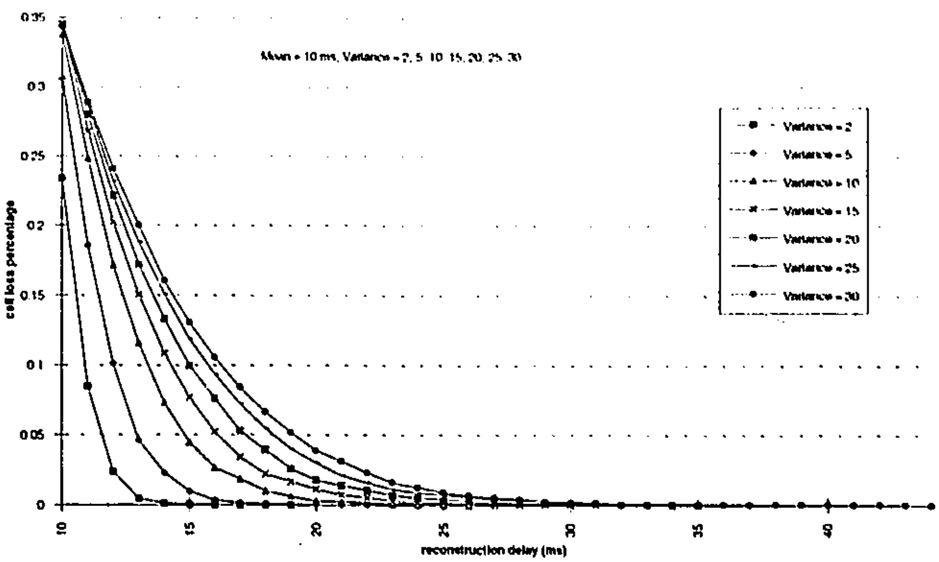


FIGURE 6.7 Relationship between cell delay variance and reconstruction delay

6.2 Echo Path Loss and Speech Quality Degradation

The objective of this experiment is two fold: a) to verify the echo emulation capability of the speech emulator; b) to investigate the impact of echoes to the speech quality in ATM network. Because of the hardware limitation, the maximum cell transmission delay that can be emulated on the emulator is about 100 ms. This is long enough to emulate the echo impact in a local area network, but it may not long enough for a long-haul network.

During the emulation, the total cell delay varies from 20 ms to 100 ms at each different echo insertion loss. Then MOS score is obtained through subjective evaluation. At the end of emulation, a set of curves is obtained to describe the relations of MOS, cell transmission delay and echo path losses.

6.2.1 Emulation Conditions

The basic configuration of the emulation is similar to the previous configuration as defined in *Section 6.1.1.1* with the following changes:

- Fixed network delay is set to 0 ms, 30 ms and 80 ms respectively
- Reconstruction delay is set to 20 ms which combines with fixed network delay to form total cell transmission delay at 20 ms, 50 ms and 100 ms respectively.
- Only 48 octets cell sized is used in the emulation
- Echo is inserted at 10, 20, 30, 40 and 50 dB respectively for each level of cell delay.
- Random cell loss rate is set to zero

Based on the configuration, total 3 x 5 = 15 sub-tests have been performed to each evaluator in order to achieve the results given in next section.

6.2.2 Emulation Results and Analysis

According to the configuration, the speech quality will only be affected by the echoes inserted into the system. It may have a very small amount cells are discarded because of the extensive cell delay, but the percentage is around 0.26% which will not have any major impact to the experiment.

The experiment results given in *Table 6.4* and *Figure 6.8* indicate that when the echo path loss is fixed, the speech quality will be degraded as total cell transmission delay increases. On the other hand, if the total cell transmission delay is fixed, the speech quality will be improved as echo path loss increases.

TABLE 6.4. Mean Opinion Score versus echo insertion and delay

Echo loss	10 dB	20 dB	30 dB	40 dB	50 dB
Total Delay	MOS	MOS	MOS	MOS	MOS
20 ms	2.56	3.52	3.76	3.95	4
50 ms	0.72	2.52	3.5	3.7	3.82
100 ms	0.01	1.2	2.95	3.4	3.56

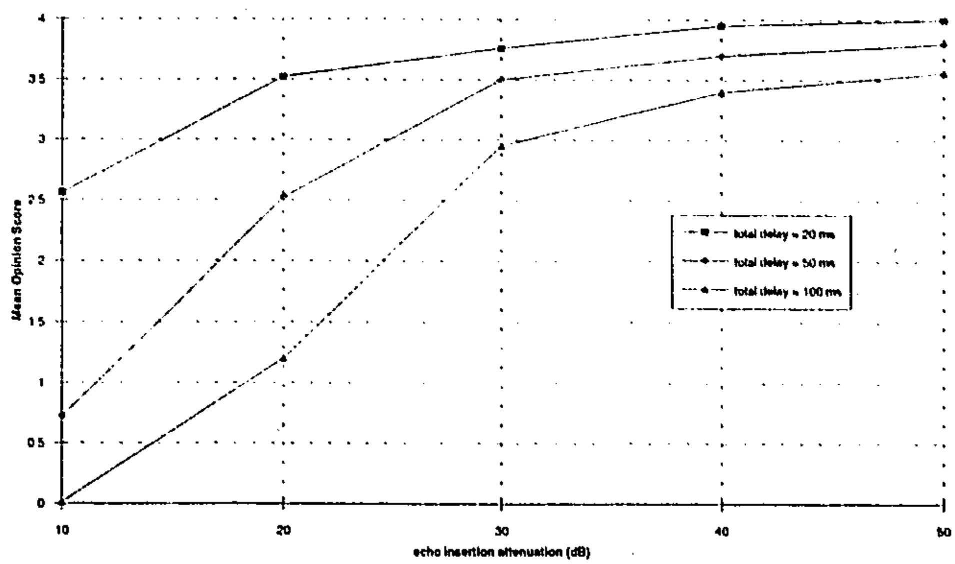


FIGURE 6.8 Mean Opinion Score on echo interference

As being discussed earlier, in digital system, as soon as the echo signal is converted into digital format, it will not be attenuated any longer. In order to reduce the impact of echoes to the speech quality in ATM network, the echo signal should be attenuated as much as possible before it is digitized. Otherwise, the digitized echo energy will go with the useful speech signal all the way down to the talker end without any further attenuation. In this case, the echo canceller has to be considered before the speech data get a cross the ATM network gateway.

6.3 PVR Algorithm Comparison

In this experiment, we will compare the performance of two types of PVRs, fixed delay compensation PVR (please refer to *Section 4.5.1*) and adaptive delay compensation PVR (please refer to *Section 4.5.2*). The basic emulation configurations are the same as the configuration defined in *Section 6.1.1.1* , but with the following changes:

- Mean cell queueing delay is kept the same, but the variance will be changed from 5 to 30 in granularity of 5
- For both fixed delay and adaptive delay PVR tests, initial reconstruction delay is set to 17 ms
- Test was only carried out based on 48 octets cell
- No echo is inserted
- Adaptation coefficient is set to 1.6.

From the experiment results given in *Table 6.5* and *Figure 6.9* , we can clearly see the performance difference between these two PVRs when cell delay variance is changed. For fixed delay compensation PVR, the performance is degraded as the cell delay variance increases. For long-haul calls, the speech cell could go through several links or switches. The network queueing delay could be changed dramatically when the network traffic condition changes. The fixed delay compensation PVR can not change its reconstruction delay T_r as the network traffic changes. As soon as the network delay exceeded the maximum compensation delay, the cell will be discarded at the PVR input. As the rate of discarded cell goes up, the service quality would go down quickly.

TABLE 6.5. Comparison between Fixed and Adaptive Delay PVRs

	fixed delay comp.		adaptive delay comp.	
Variance	MOS	Qop	MOS	Qop(dB)
5	4.03	36.65	4.02	36.6
10	3.48	27.7	4.01	35.5
15	2.9	24.3	3.98	34.8
20	2.33	19.9	3.89	32
25	1.95	17.3	3.87	31.98
30	1.8	16.2	3.87	31.98

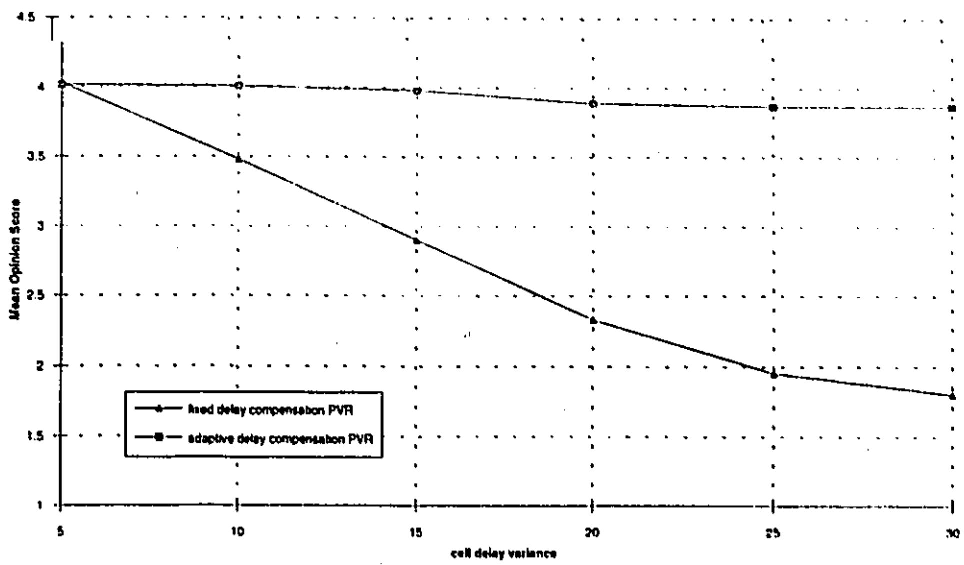


FIGURE 6.9 Mean Opinion Score between Fixed Delay and Adaptive Delay PVRs

In the case of adaptive delay compensation PVR, from *Figure 6.9* we can clearly see that the speech quality almost keeps the same as cell delay variance increases, because the adaptive PVR algorithm is automatically adjusting the reconstruction delay by compressing or expanding the silence period to compensate the network queueing delay variance changes. For a long-haul communications network, the adaptive PVR is definitely a way to go in the future.

6.4 Summary

This chapter presents an overview of ATM speech communication emulation and experiment results analysis. For Poisson distributed cell losses, 3% cell loss rate was found to be the limit that human hearing can tolerate for PCM waveform coder. Echo impact to the speech quality in ATM network has been proved through the emulation. The findings through the experiments are very useful for further study on the topic. It will also be helpful for the industry to get better understanding of ATM speech communications.

CHAPTER 7 Conclusions and Future Research Topics

7.1 Summary and Conclusions

In this thesis we have investigated the issues related to the transport of speech signals in ATM networks and the effects of these issues to the quality of speech communications. These issues include: cell loss rate, cell packetization delay, cell transmission delay, cell network queueing delay and distribution, ATM packet voice receiving algorithm and the evaluation of ATM speech communication.

Based on the results of the investigation, a prototype ATM speech emulator has been designed and implemented on a personal computer and DSP5600 ADM system. The emulator provides the capability to support one full duplex voice channel real time emulation. The speech quality could be subjectively evaluated based on actual telephone conversations over the emulator.

A set of emulations has been performed, through subjective evaluation, on the prototype emulator which confirms the correct functionality and the accuracy of the emulator. The emulation study has resulted in a number of important findings.

In ATM networks, speech quality will be degraded as total cell missing rate increases. When cell missing rate reaches 3% in total, which includes both the cell loss by network transmission error and the cell loss by excessive cell arrival delay, the voice quality will drop to a level (MOS = 2.5) which can not be accepted in public telephone network. These results, alone with Kitawaki [12], indicate that 3% combined cell loss rate will be upper bounded for ATM speech communications, no matter which error source dominates the 3% cell loss rate. However, this result is only suitable for purely 64 Kbits PCM bit stream without any other protective algorithm being used, such as shared cell payload or missing cell recovery.

The speech quality will be degraded when cell size increases. Comparing 32 octets cell with 48 octets cells, the speech quality of 32 octets cell will be a little better than 48 octets cell. When cell loss rate is lower than 2.5%, the speech quality between these two options are very close and when the cell loss rate is higher 4%, the speech quality with longer cell has more obvious degradation than shorter cell does. This result indicates that within acceptable cell loss range, both European and American proposed standard will achieve reasonable performance on speech quality. European proposal is a little bit better on speech quality and North American proposal is a little bit better on channel utilization.

Echo is still a considerable factor which may potentially degrade the speech quality in ATM networks. It is not a problem in a full digital network if acoustic echoes can be completely suppressed before they are transmitted. If ATM network is integrated with existing system and telephone equipment, echo cancellation technique has to be considered for the long-haul communications unless the network infrastructure can keep the total cell delay relatively small so that does not cause any obvious echo effect. For example, if total echo loss is 20 dB, the system has to keep the overall network delay less than 30 ms in order to keep reasonable speech quality (MOS >2.6).

Two PVR algorithms have been compared through emulation, the results show that adaptive delay compensation PVR is much better than fixed delay compensation PVR when the network delay variance dynamically changes. When cell delay variance changed from 5 to 30, the speech quality only dropped about 0.15 on MOS scale. This indicate that adaptive PVR algorithm has the better capability to handle large cell delay fluctuation caused by bursty network traffic which is a typical traffic pattern in ATM network. For a long-haul voice communications in ATM network, adaptive PVR algorithm should be considered.

7.2 Future Research in ATM Speech Communications

Because of hardware and time limitation, many useful functions and features on speech communication emulations could not be implemented in this prototype speech emulator, such as the capability of using different CODECs[5], cell packetization techniques[49], more PVR algorithms[51], the missing cell recovery mechanisms and the background noise estimation/insertion techniques[37]. Some of them are more straightforward implementation efforts which only requires more hardware capabilities like more computing power and memory and some of them do need further studies and simulations. These will be the main directions of further study on ATM speech communications.

In current implementations, the speech emulator only supports A-law and μ-law PCM CODECs. In fact, the 32 kbits ADPCM and 16 kbits LPC CODECs have become mature and been deployed in both public and private network for several years. How do the speech samples from these CODECs fit into the ATM cells? And how is the cell assembled and disassembled? How do the cell packetization techniques for these CODECs impact the speech quality? It is worthwhile to get these CODECs implemented and try them out on the emulator.

ATM cell packetization and missing cell recovery are closely related two areas. The study on ATM packetization will definitely have the impact to the cell recovery techniques. In general, the research on these topics can certainly benefit on following two aspects: a) reduce total cell delay by reducing cell packetization delay; b) reduce the difficulties on missing cell recovery. Normally, if we want to reduce the error recovery difficulties, we have to reduce the information lost duration (less samples together). To achieve the objective, the speech samples have to be spread out in several cells (partially filled cell), which will definitely reduce the cell packetization delay. But if the payload of cells are not shared with other speech sources, the channel utilization will be reduced significantly. The study for a better algorithm is always required to meet the requirements of short packetization delay and easy missing cell recovery. If a shared cell payload strategy is considered, the network administration overhead has to be studied carefully.

Packet voice receiving algorithm is the core of ATM voice receiver. The speech quality of ATM voice communications heavily relies on the performance of the packet voice receiving algorithm. For the prototype speech emulator, only a simplified adaptive receiving algorithm was implemented with adaptive coefficient α = 1.6. The cell delay variance and cell discarding rate are not considered in the implementation. In order to derive a more accurate adaptive receiving algorithm, the relations between T and *Var*, P_c and D_{ma} in *EQ. 7 - 1* need to be carefully studied. Reconstruction delay T is a function of *Var*, *Pc* and D_{ma}, where *Var* is measured queueing delay variance in talkspurt $(i - 1)$, P_c is cell discarding rate in talkspurt $(i - 1)$ at PVR and D_{ma} is maximum cell arriving delay after target play-out time in talkspurt $(i - 1)$. i is the talkspurt number in a call, $i = 1, 2, \ldots n$. α is the safety coefficient which is direct proportional to Var and P_c.

$$T_i = f\{Var_{i-1}, Pc_{i-1}, \alpha_i(Var_{i-1}, Pc_{i-1}) \cdot D_{ma}\} \quad \text{(EQ 7 - 1)}$$

Further study on this relation and coefficient should result in a more accurate reconstruction delay estimation when the network traffic condition varies dynamically.

ATM is a very promising technology for future telecommunication networks. In the past years, tremendous research on ATM network have been conducted and the progress is amazing. The pioneer ATM switch and network products have already been developed based on the results of earlier studies and TIA and CCITT preliminary standards. The findings of this study has further proved that speech transmission over ATM network is feasible and will definitely reach the similar quality level as existing telephone networks in the near future.

REFERENCES:

[1] Russell G. DeWitt, "ISDN Symposia: A Historical Overview," IEEE Communications Magazine, Vol. 28, No.4, April 1990.

[2] CCITT Recommendation I.121, "Broadband Aspects of ISDN", CCITT Blue Book, Volume III, Fascicle III.7, Geneva 1989.

[3] Bell Laboratories, "Transmission System for Communications", Fifth Edition, Bell Telephone Laboratories Inc. 1982.

[4] Freeman, Roger L, "Telecommunication Transmission Handbook", 3rd Edition, New York, Wley, 1991.

[5] Nobuhiko Kitawaki, Hiromi Nagabuchi, Masahiro Taka and Keno Takahashi, "Speech Coding Technology for ATM Networks", IEEE Communications Magazine, Vol.28, No. 1, pp. 21-27, January 1990.

[6] N.S. Jayant, "High-Quality Coding of Telephone Speech and Wideband Audio", IEEE Communications Magazine, Vol. 28, No.1, pp. 10-20, January 1990

[7] T. C. Spang, "Loss-Noise-Echo Study of the Direct Distance Dialing Network," Bell System Technical Journal, vol.55, pp.1-36, Jan. 1976.

[8] E. W. Holman and V. P. Suhocki, "A New Echo Suppressor," Bell Laboratories Record, April 1966.

[9] R. C. Drechsler, D. C. Koehler, and R. D. Royer, "Suppressing Echoes Digital Style," Bell Laboratories Record, vol. 56, pp. 142-146, June 1978.

[10] R. C. Drechsler, "Echo Suppressor Terminal for No. 4 ESS," Proc. IEEE ICC' 76, Vol. III, pp. 36-9, 36-12, Philadelphia, PA, June 1976.

[11] M.M. Sondhi, "An Adaptive Echo Canceller," Bell System Tech. J., vol. 46, pp. 497-511, Mar. 1967.

[12] N. Kitawaki and H. Nagabuchi, "Quality Assessment of Speech Coding and Speech Synthesis Systems," IEEE Commun. Mag., vol.26, no. 10, Oct. 1988.

[13] CCITT, Recommendation P.81, "Modulated Noise Reference Unit (MNRU)," Blue Book, vol. V, 1988.

[14] H. Nagabuchi, N. Kitawaki, "Evaluation of Coded Speech Quality Degraded by Cell Loss in ATM Network," Electronics and Communications in Japan, Part 3, Vol. 75 Iss: 9, pp 14-24, Sept. 1992

[15] Jaime Jungok Bae and Tatsuya Suda, "Survey of Network Congestion Control Protocol," Proceeding of GLOBECOM'90, pp. 3001.1 - 300.1.6, San Diego, CA, Dec. 1990.

[16] G. Gallassi, G. Rigolio and L. Fratta, "ATM: Bandwidth Assignment and Bandwidth Enforcement Policies," Proc. IEEE GLOBECOM'89, pp.49.6. 1-49.6.6, Dallas, TX Dec 1989.

[17] M. Hirano and N. Watanabe, "Characteristics of a Cell Multiplexer for Bursty ATM Traffic," Proc. IEEE ICC '89, pp.13.2.1-13.2.5, Boston, June 1989.

[18] S. Akhtar, "Congestion Control in a Fast Packet Switching Network," Master's Thesis, Washington University, December 1987.

[19] J. Y. Hui and E. Arthurs, "A Broadband Packet Switch for Integrated Transport," IEEE J. Select. Areas Commun., SAC-5, No.8, pp.1264-1273, October 1987.

[20] K. Sriram and W. Whitt, "Characterizing Superposition Arrival Processes in packet Multiplexers for Voice and Data," IEEE J. Select. Areas Commun., SAC-4, No.6, pp.833-846, September 1986.

[21] A. E. Eckberg, "Generalized Peakedness of Teletraffic Processes," Proc. 10th International Teletraffic Congress, Montreal, pp. 3-7, June 1983.

[22] T. Kamitake and T. Suda, "Evaluation of an Admission Control Scheme for an ATM Network Considering Fluctuations in Cell Loss Rate," Proc. IEEE GLOBECOM'89, pp.49.4.1-49.4.7, Dallas, TX Dec 1989.

[23] S-Q. Li, "Study of Information Loss in Packet Voice Systems," IEEE Trans. Commun., Vol. 37, No. 11, pp. 1192- 1202, November 1989.

[24] J. S. Turner, "New Directions in Communications (or Which way to the Information Age?)," IEEE Commun. Magazine, Vol.25, No.10, pp.8-15, October 1986.

[25] I. Cidon and I. S. Gopal, "PARIS: An Approach to Integrated High-Speed Private Networks," Int. J. Digital & Analog Cabled Systems, Vol.1, No.2, pp.77-86, April-June, 1988.

[26] Giulio Barberis, Daniele Pazzaglia, "Analysis and Optimal Design of a Packet-Voice Receiver", IEEE Transactions on Communications, Vol.COM-28, No.2, pp. 217 - 227, February, 1980.

[27] Clifford J. Weinstein and W. Forgie, "Experience with Speech Communication in Packet Network", IEEE Journal on Selected Areas in Communications, Vol.SAC-1, NO.6, pp.963 - 980, December 1983.

[28] E. Tsutsui, H. Fukuda, S. Nojima, M. Hashimoto and T. Hasebe, "Evaluation of the Maximum Packet Delay of a High Speed Packet Switch", Proc. IEEE GLOBECOM'87, pp47.5.1- 47.5.5, Tokyo, Japan, Nov 1987.

[29] G. Ramamurthy and R.S.Dighe, "Performance Analysis of Circuit-Switched Traffic over a Packet Network", Proc. IEEE ICC'90, pp.310.51-310.5.7. New York, June 1990.

[30] Steven E. Minzer, "Broadband ISDN and Asynchronous Transfer Mode (ATM)," IEEE Communications Magazine, Vol.27, No. 9, pp.17-24, September 1989.

[31] Toshiharu Hasegawa, "Performance Evaluation of a Packetized Voice System Simulation Study", IEEE Transactions on Communications, Vol. COM-32, NO.1, January 1984.

[32] Warren A. Montgomery, "Techniques for Packet Voice Synchronization", IEEE Journal on Selected Areas in Communications, Vol. SAC-1, NO.6, pp. 1022-1028, December 1983.

[33] J. G. Gruber, "Delay related issues in integrated voice and data networks: A case study," IEEE Transactions on Communications, vol. 29, pp. 786-800, June 1981

[34] E. T. Klemmer, "Subjective evaluation of delay and echo suppressors in telephone communications," Bell System Tech. J., vol.42, Nov. 1963

[35] J. W. Forgie, "Network speech system implications of packetized speech", Massachusetts Inst. Technol., Lincoln Lab., Tech. Annul. Rep. Sept. 30, 1976

[36] William O. Covington, "Voice Transport on an ATM Broadband Network", Proceeding of ICC'89, pp.1921-1925, Boston, June 1989.

[37] Takao Suzuki, Osamu Noguchi, "A New Processing Scheme for ATM Switching Systems", Proceeding of ICC'89, pp.1515-1519, Boston, June 1989.

[38] Masayuki Murata and Yuki Oie, "Analysis of a Discrete-Time Single-Server Queue with Bursty Inputs for Traffic Control in ATM Networks", IEEE Journal on Select Areas in Communications. Vol.8 No.3, April 1990

[39] CCITT Study Group XVIII Draft Report, "COM XVIII-R55(B)", 1988

[40] Giulio Barberis, Mario Calabrese, "Coded Speech in Packet- Switched Networks: Models and Experiments", IEEE Journal on Select Areas in Communications. Vol.SAC-1, No.6, December 1983

[41] J.F. lynch Jr., J.G. Josenhans and R.E. Crochiere, "Speech/Silence Segmentation for Real-Time Coding via Rule based adaptive Endpoint Detection", IEEE ICASSP'87 Proceedings, pp. 1348- 1351, Dallas, Texas, Apr. 1987.

[42] Y.Shoji, O. Noguchi and T.Suzuki, "Development of High Performance DCMS with 3-bit and 4-bit Coding ADPCM," Proceedings of IEEE ICC'88, pp. 1598-1602, Philadelphia, PA, June 1988.

[43] D.J.Goodman, O.G.Jaffe, G.B.lotkhart and W.C. Wong, "Waveform Substitution Techniques for Recovering Missing Speech Segments in Packet Voice Communications", IEEE ICASSP'86 Proceedings, pp.105-108, Tokyo, Japan, Apr. 1986.

[44] L. R.Rabiner and R.W. Schafer, "Digital Processing of Speech Signals", Prentice-Hall, 1978.

[45] N. Kitawaki, M. Honda and K. Itoh, "Speech Quality Assessment Methods for Speech Coding Systems," IEEE Communications Magazine, Vol. 22, pp.26-33, Oct. 1984.

[46] Hiromi Nagabuchi, Akira Takahashi and Nobuhiko Kitawaki, "Speech Quality Degraded by Cell Loss in ATM Networks", NTT Review Vol.4, July 1992

[47] William O. Covington, Jr. Micheal E. Vitella, "Voice Transport on an ATM Broadband Network," Proceeding of GLOBECOM'88, pp.53.4.1-53.4.5, Hollywood, FL, Nov. 1988.

[48] T. Yamazaki, Hiroshi Irii, "Objective Estimation of Speech Quality Degradation Due to Transmission Error", Proc. IEEE SuperComm/ICC'92, pp.66-70, vol.I, Chicago, June 1992

[49] William O. Covington, Jr. Micheal E. Vitella, "Voice Transport on an ATM Broadband Network," Proceeding of GLOBECOM'88, pp.53.4.1 - 53.4.5, Hollywood, FL, Nov. 1988.

[50] M M Lara-Barron and G B Lockhart, "Speech Quality Improvement in ATM Networks Using an Alternating-Cell Discarding Strategy", Proceeding of Third IEE Conference on Telecommunications, pp.338-342, Edinburgh, UK, 1991

[51] Warren A. Montgomery, "Techniques for Packet Voice Synchronization", IEEE Journal on Selected Areas in Communications, Vol. SAC-1, NO.6, December 1983

[52] Annie Gravey and Pierre Boyer, "Cell Delay Variation Specification in ATM Networks," Modelling and Performance Evaluation of ATM Technology (C-15), Elsevier Science Publishers B. V. (North-Holland), 1993 IFIP

[53] Brady, P. T., "A Statistical Analysis of ON-OFF Pattern in 16 Conversations," BSTJ, pp.73-91, Jan 1968

APPENDIX I. Audio Interface Circuit

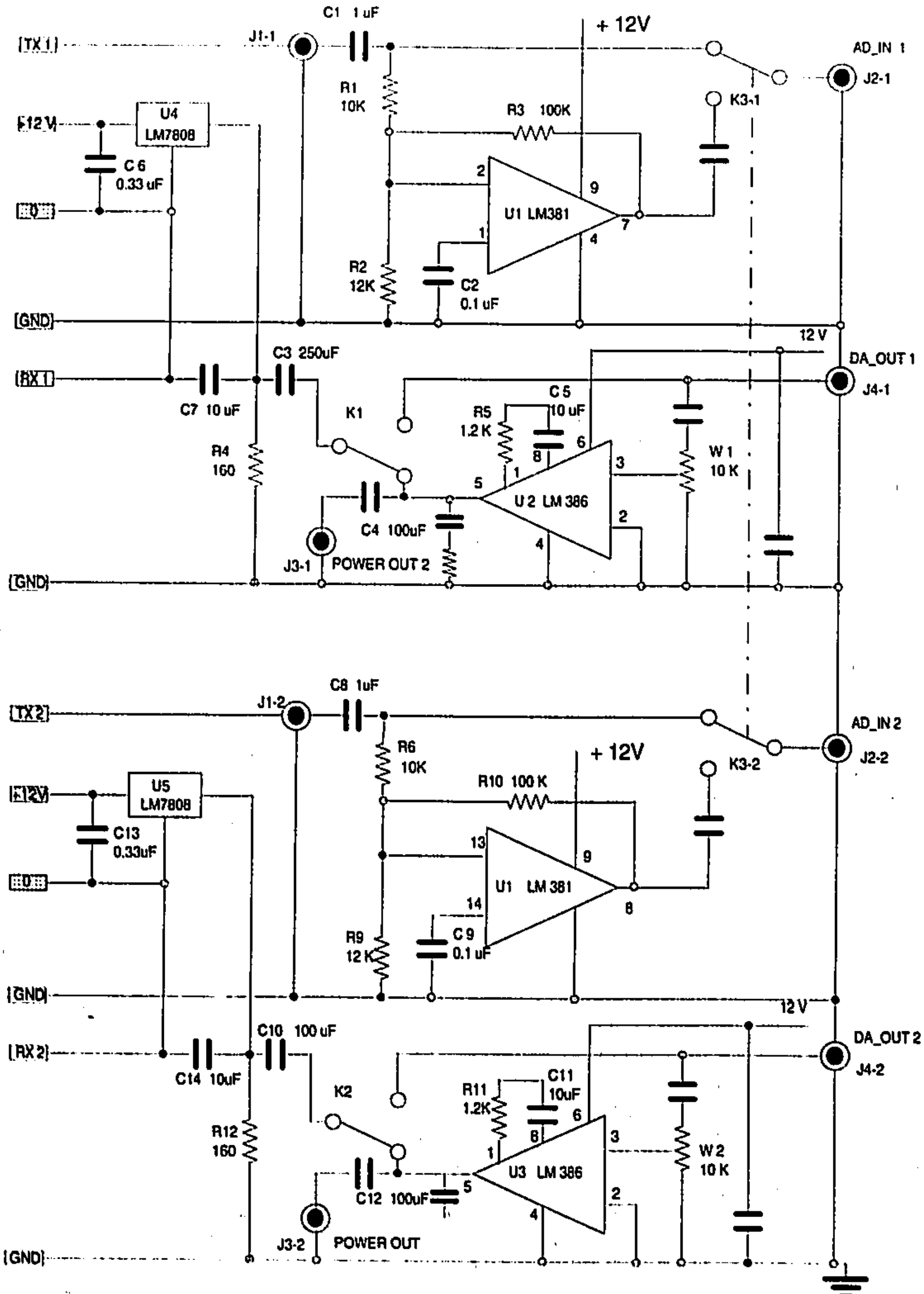


FIGURE A.1 Circuit of Audio Interface

APPENDIX II. Cell Delay Distribution Table

TABLE A.1. Cell Delay Distribution at Mean = 10 ms, Variance= 10

Cell Queueing Delay (ms)	Frequency	Cumulative%
0	0	0.00%
1	1	0.02%
2	2	0.06%
3	33	0.72%
4	110	2.92%
5	280	8.52%
6	445	17.42%
7	595	29.32%
8	647	42.26%
9	629	54.84%
10	570	66.24%
11	446	75.16%
12	385	82.86%
13	279	88.44%
14	213	92.70%
15	142	95.54%
16	92	97.38%
17	39	98.16%
18	42	99.00%
19	19	99.38%
20	18	99.74%
21	1	99.76%
22	5	99.86%
23	3	99.92%
24	1	99.94%
25	2	99.98%
26	1	100.00%
27	0	100.00%
28	0	100.00%
29	0	100.00%
30	0	100.00%

Note: This table is only an example to explain why 18 ms compensation delay is selected for the experiment.